PRIMER ON DATA ANALYSIS

TABLE OF CONTENTS

Introduction

The basic purpose of data analysis is to summarize with a few parameters the information in a data set which may consist of many measurements. One rarely publishes raw data but rather the conclusions implied by the parameters calculated from the data set. This set of notes therefore concerns itself with describing what are generally considered useful parameters to extract from data.

What does one observe about the process of acquiring data? In many cases repetitive measurement of a supposedly precisely defined quantity yields not a single precise value but a scattered set of measurements. As more and more data is accumulated one usually finds that the "middle," of the distribution of measurements becomes clearer, and that small deviations from the "middle," are much more prevalent than large ones. This implies that a distribution like that of Figure (1) is a better model for a typical data set than Figure (2). Figure (2) implies that large deviations are just as likely as small ones; a result not usually seen.

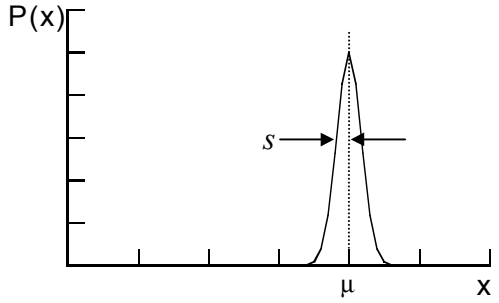


Figure (1)

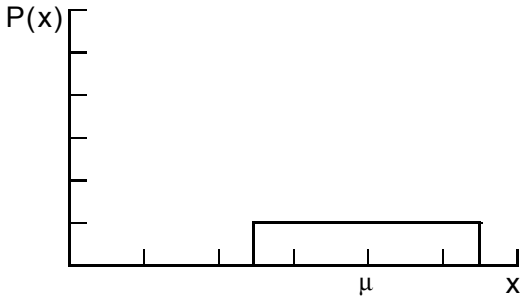


Figure (2)

Looking at Figure (1) it is clear that a parameter representing the "middle" and one representing the width of the distribution would be useful descriptors of a data set. There are several possible ways of defining the "middle" or central value of the data set with the simplest and most generally used being the familiar arithmetical average.

For a set of N measurements $x_1\ x_2, x_3, \ldots, x_N$

$$\mu \equiv \frac{1}{N}\left(x_1 + x_2 + x_3 + \ldots + x_N\right) \equiv \frac{1}{N}\sum_i^N x_i \tag{1}$$

A commonly used width descriptor, s^2, calculates the average of the square of the scatter (x_i - m) about m. Squaring allows the negative terms to contribute without the possibility of the scatter averaging out to zero. The quantity s, is the standard deviation.

$$s^2 \equiv \frac{1}{N}\left[\left(x_1 - \mu\right)^2 + \left(x_2 - \mu\right)^2 + \left(x_3 - \mu\right)^2 + \ldots + \left(x_N - \mu\right)^2\right] \tag{2}$$

Equation (2) relies on a parameter previously obtained from the data set, using equation (1). To give an unbiased estimate of the standard deviation *s*, equation (2) is modified to give one less degree of freedom (*N*-1), and is usually stated as

$$s^2 \equiv \frac{1}{N-1}\sum_{1}^{N}(x_i - \mu)^2 \tag{3}$$

Take for example five measurements of a precession period T in seconds

$T_1 = 59.35 \quad T_2 = 60.23 \quad T_3 = 58.76 \quad T_4 = 59.83 \quad T_5 = 58.98$ sec

$\mu = (59.35 + 60.23 + \ldots) / 5 = 297.15 / 5 = 59.43$ sec
$s^2 = [\,(59.35 - 59.43)^2 + (60.23 - 59.43)^2 + \ldots\,] / 4 = 0.36$ sec^2

Thus this distribution can be summarized by stating μ and *s*, i.e.,

$$\mu = 59.43 \text{ sec}, \; s = 0.6 \text{ sec}$$

Note that if *s* is a good estimate of the width it would not change significantly after 20 measurements but our estimate of μ would be much better. Thus the best estimate of μ should improve with the number of measurements, leading to

$$\boxed{\mu \pm s/\sqrt{N}} \tag{4}$$

which in this case gives

$\underline{\mu = 59.43 \pm 0.27 \text{ sec}}$ as the best estimate.

Although the mean, μ, and the standard deviation of the mean, $s/\sqrt{N}$, may appear to have been rather arbitrarily adopted for the purpose, no other expressions for the "best single value" and the "scatter" about that value, provide more consistent, more accurate, or more convenient measures. In addition, these quantities are firmly rooted in an extensive and fundamental theory of statistics and data analysis, with which the student will gradually become more familiar.

The Primer deals with the analysis of <u>random uncertainties</u>, but in any experiment the possibility of <u>systematic errors</u> is always present . Systematic errors occur in a variety of forms due either to equipment design and calibration or faulty technique on the part of the experimenter. Typical equipment problems include backlash in screw threads, an offset in the zero set of an analog meter, etc. Typical systematic errors related to the experimenter's technique include parallax, the failure to place the eye correctly when sighting on two marks, bias, such as a tendency to always read a value on the high side, etc. There is no mathematics to account for systematic errors it is therefore up to the experimenter to stay alert to the possibilities and work to eliminate or, at the very least, minimize them.

Finally a word about precision and accuracy. Precision is associated with the size of the random uncertainties and accuracy is associated with systematic errors. As an example, m is the

best known value for some quantity which is measured by two different techniques resulting in $A \pm s_A$ and $B \pm s_B$.

Result *A* has higher precision but is clearly suffering from a lack of accuracy, whereas result *B*, though lower in precision, has better accuracy and is within 2 standard deviations of m (a typical good result).

Mean and Standard Deviation

A set of measurements of a given quantity is always a sample from some parent distribution. Any given measurement is only an approximation to the parameter you are seeking. On the average, however, one expects the measurements to be randomly distributed around the best value, and that the experimentally determined parameters will approach the parent parameters as the number of measurements, $N \to \infty$.

The parent distribution which best describes the randomly distributed experimental values is usually the "Gaussian" or "Normal" distribution. The probability $P(x)$ of observing a value between x and $x + dx$ is described by the Gaussian probability density

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad (1)$$

where μ is the mean value of x and σ is the standard deviation. This distribution has the familiar shape shown in Figure 1 which is a plot of $P(x) = e^{-(x-\mu)^2/2\sigma^2}$

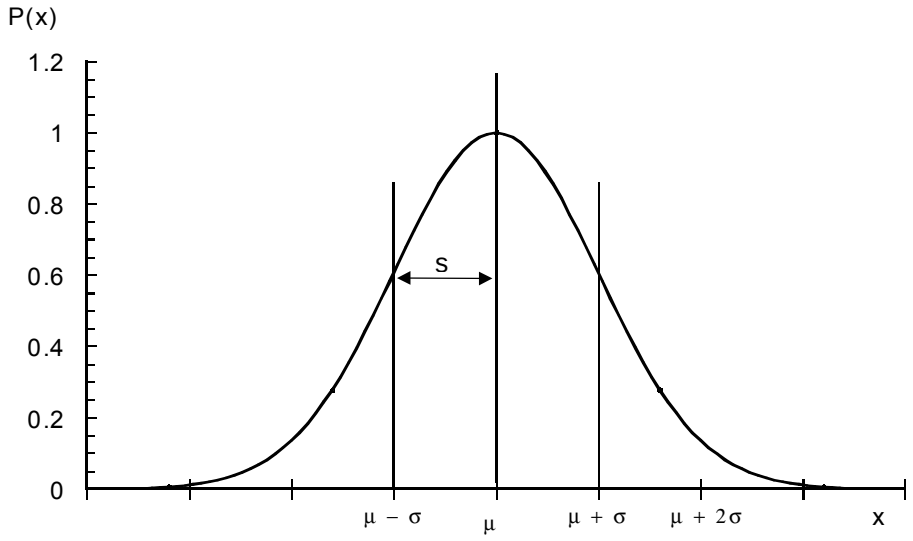


Figure 1

The distribution of Eq. (1) is normalized so that the area under the curve is unity. The probability that a single observation of x will be within one σ of μ is equal to the fraction of the area under the curve between $\mu - \sigma$ and $\mu + \sigma$ which is approximately 68%.

$$Area = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-\sigma}^{\mu+\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

let $t = \dfrac{x-\mu}{\sigma}$ then $dx = \sigma\, dt$

$$Area = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-t^2/2}\, dt = 0.682 \qquad \{\text{See any book on Integrals}\}$$

Also at $x = \mu$ $P(x) = 1/(\sigma\sqrt{2\pi}) = A$ and at $x = \mu + \sigma$, $P(x) = Ae^{-0.5} \cong 0.6A$. Thus σ is roughly the half width at the half height of a Gaussian curve. (See Figure 1)

The applicability of the Gaussian distribution to a given set of data can be tested in a number of ways. Clearly a strict test cannot be made unless a great many measurements are available, since it is the measurements that lie several standard deviations from the mean that are critical to the test. Thus if only two observations are made, nothing can be said about the applicability of the Gaussian distribution. However, if 100 observations are available, a very sensitive test may be made by comparing the number of deviations lying outside a given value with the number expected from the Gaussian distribution. A rough test can be made for as few as ten data points, by noting whether about a third of them lie outside one standard deviation from the mean.

Mean μ of the parent population

If the probability of observing any one measurement in the parent population is $P(x_i)$ given by equation (1), then the probability of observing a set of measurements is the product of the individual probabilities.

$$P_s = \prod_{i=1}^{N} P(x_i)$$

This gives

$$P_s = \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^N \exp\left[-\frac{\sum(x-\mu)^2}{2\sigma^2}\right] \tag{2}$$

Looking at Figure 1 suggests that the best value for m is obtained by maximizing P_s, or equivalently minimizing the sum in the exponential

$$\frac{d}{d\mu}\frac{\sum(x_i-\mu)^2}{2\sigma^2} = 0$$

Noting that $\frac{d}{d\mu}\sum = \sum\frac{d}{d\mu}$

gives $\Sigma(x_i - m) = \Sigma x_i - \Sigma m = \Sigma x_i - Nm = 0$

thus

$$\boxed{\mu = \frac{1}{N}\sum x_i} \tag{3}$$

Equation (3) says that if all the individual measurements are equally good (i.e. the uncertainty s is the same for each) then the best value for m is just the simple unweighted average of all the measurements. The implication here is that all the measurements were from the same parent population described by a single distribution characterized by a single value for m and s.

In the general case where individual measurements may have different sigmas (uncertainties) associated with them equation (3) becomes

$$\frac{d}{d\mu}\sum\frac{(x_i-\mu)^2}{2\sigma_i^2} = 0$$

$$\boxed{\mu = \frac{\sum\left(x_i/\sigma_i^2\right)}{\sum\left(1/\sigma_i^2\right)}} \tag{4}$$

Equation (4) gives m in terms of a weighted average of the measurements. The implication

is that the various measurements were from different parent populations, all having the same m but different widths s.

Standard Deviation

The commonly used width parameter s^2, usually refered to as the variance, mentioned in the INTRODUCTION is generally easy to calculate. So the question is how is it related to the σ of the Gaussian distribution? The variance of a distribution, $f(x)$, is defined as

$$s^2 \equiv \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

Which for a Gaussian distribution gives $$s^2 \equiv \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$$

Making a change of variable, $z = (x - m)/s \div 2$, gives an integral which can be evaluated by parts

giving $$s^2 \equiv \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2} dz = \frac{2\sigma^2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = \sigma^2$$

Thus it is seen that s of the Gaussian distribution can be estimated by

$$s^2 = \lim_{N\to\infty}\left\{\frac{1}{N}\sum (x_i - \mu)^2\right\} \cong \sigma^2 \qquad (5)$$

Practical Considerations

In reality you will only make a finite number of measurements. How does this fact modify the above equations? If N is finite then

$$\bar{x} = \frac{1}{N}\sum x_i \cong \mu \qquad (6)$$

$$\boxed{s^2 = \frac{1}{N-1}\sum (x_i - \bar{x})^2 \cong \sigma^2} \qquad (7)$$

Where the parent standard deviation s, is estimated by ***s***, the sample standard deviation. The quantity s^2, also known as the variance, can be written in the form

$$s^2 = \frac{1}{N-1}\left[\sum x_i^2 - \frac{1}{N}\left(\sum x_i\right)^2\right]$$

The N-1 term is the number of degrees of freedom, which is defined as the number of observations minus the number of terms needed in a theoretical description of the data.

For a point x = a	Degrees of freedom is N - 1
For a straight line y = a + bx	Degrees of freedom is N - 2
For a parabola $y = a + bx + cx^2$	Degrees of freedom is N - 3

The N -1 term in equation (7) also avoids the erroneous result $s^2 = 0$, in the case where only one measurement is made. Another way to look at the N -1 term is that having used the data to calculate $\bar{x}$ there is one less degree of freedom in the data set when calculating the next parameter.

Note: The variance s^2 of equation (7) is an estimate of the average uncertainty of the individual data points, x_i, as well as an estimate of the width of the parent distribution.

Sigma of the Mean

After N measurements the mean m, is known to a smaller uncertainty than the s of the distribution from which the N measurements were drawn.

In general the s^2 of any quantity is equal to the sum of the s_i^2 of the individual data points times the square of the effects of that variable on the result.

(See eq.(4) pg. 2-2)

$$\boxed{\sigma_\mu^2 = \sum \sigma_i^2 \left(\frac{\partial \mu}{\partial x_i}\right)^2} \qquad (8)$$

Now $\mu = \frac{1}{N}\sum x_i$ thus $\sigma_\mu^2 = \sum \sigma_i^2 \left(\frac{1}{N}\right)^2$

If all the s_i are equal, then $s_i = s$ giving

$$\sigma_\mu^2 = \sigma^2 \sum_1^N \left(\frac{1}{N}\right)^2 = \frac{\sigma^2}{N} \qquad (9)$$

Thus the mean should be stated as

$$\boxed{\mu \cong \bar{x} \pm \frac{\sigma}{\sqrt{N}}}$$

If the uncertainties s_i on the individual data points are not equal we will show later on that the mean m and s_m are equal to (See page 1-8)

$$\boxed{\mu \cong \bar{x} = \frac{\sum\left(x_i / \sigma_i^2\right)}{\sum\left(1/\sigma_i^2\right)}} \tag{10}$$

$$\boxed{\sigma_\mu^2 = \frac{1}{\sum\left(1/\sigma_i^2\right)}} \tag{11}$$

Another way to derive σ_μ

What follows depends on ideas developed in the next chapter on propagation of uncertainties.

A quantity is measured N times ($x_1, x_2, x_3, \ldots$). The total is x_T where

$x_T = \sum x_i$ and $s_T = (s_1^2 + s_2^2 + s_3^2 + \ldots)^{1/2}$ (See page 2-3)

The average value $\bar{x}$ is $\bar{x} = \frac{x_T}{N}$

And $\sigma_{\bar{x}}$ is (see page 2-2 equation (1))

$$\sigma_{\bar{x}} = \frac{\partial \bar{x}}{\partial x_T}\sigma_T = \frac{\sigma_T}{N}$$

If all the measurements have the same uncertainty s, then $\sigma_{\bar{x}}$ is

$$\sigma_{\bar{x}} = \frac{\sigma_T}{N} = \frac{\sigma\sqrt{N}}{N} = \frac{\sigma}{\sqrt{N}}$$

Example 1

Ten periods of a torsion pendulum are measured five times giving

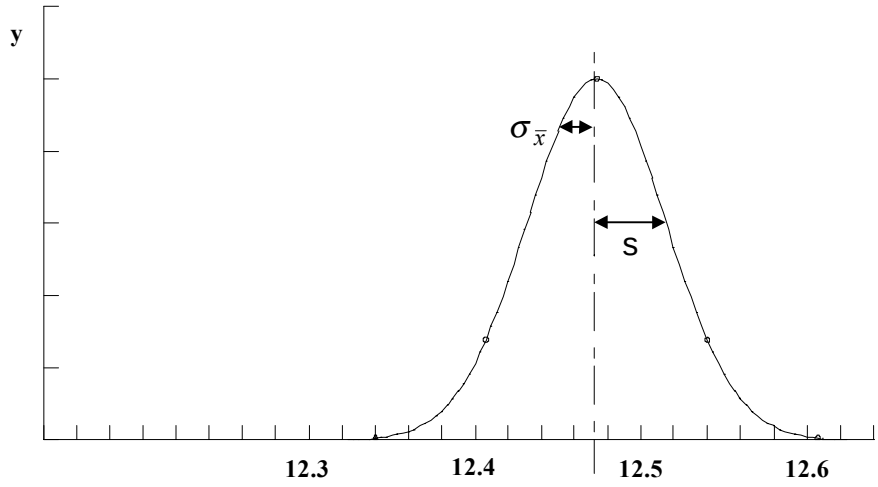
$10T$ = 12.52, 12.51, 12.45, 12.42, 12.46 sec.

Using equations (6) (7) and (9)

$\bar{x} = \overline{10T} = 12.47$ $\quad s$ @ s = 0.042 and $\quad \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{5}} = 0.02$

Assuming a Gaussian parent distribution from which these presumably randomly distributed measurements were drawn

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp - \left[\frac{\left(10T - \overline{10T}\right)^2}{2\sigma^2} \right]$$



Thus the 10 period average is recorded as

$$\overline{10T} = 12.47 \pm 0.02 \,\text{sec}$$

and

$$\overline{T} = 1.247 \pm 0.002 \,\text{sec}$$

Example 2 Weighted Average

A quantity *b*, is measured several times, each measurement having a different uncertainty. What is the weighted average value *b*, and its uncertainty?

$$b_1 = 7.4 \pm 0.3$$
$$b_2 = 7.9 \pm 0.4$$
$$b_3 = 7.5 \pm 0.2$$

$$\bar{b} = \frac{\sum b_i / \sigma_i^2}{\sum 1/\sigma_i^2} = \frac{\dfrac{7.4}{(0.3)^2} + \dfrac{7.9}{(0.4)^2} + \ldots}{1/(0.3)^2 + 1/(0.4)^2 + \ldots} = 7.5 \qquad (\sigma_{\bar{b}})^2 = \frac{1}{\sum 1/\sigma_i^2} = 0.02$$

Thus $\boxed{\bar{b} = 7.5 \pm 0.1}$ (An unweighted average gives $b = 7.6$)

Derivation of σ_μ^2 in the Weighted Case

Equation (8) states

$$\sigma_\mu^2 = \sum \sigma_i^2 \left(\partial\mu / \partial x_i \right)^2 \qquad \text{where} \quad \mu = \frac{\sum x_i / \sigma_i^2}{\sum 1/\sigma_i^2}$$

Each term in the sum of equation (8) is generated by taking the partial derivative of m with respect to a particular x_i, namely $i = j$.

Thus $$\frac{\partial \mu}{\partial x_j} = \frac{1/\sigma_j^2}{\sum 1/\sigma_i^2}$$

Consider the first few terms of the sum (j = 1, 2, ...)

$$\sigma_\mu^2 = \left[\frac{1}{\sum 1/\sigma_i^2} \right]^2 \left\{ \sigma_1^2 \frac{1}{\sigma_1^4} + \sigma_2^2 \frac{1}{\sigma_2^4} + \cdots \right\}$$

Giving $$\boxed{\sigma_\mu^2 = \frac{1}{\sum 1/\sigma_i^2}}$$

A Comment on Averaging Data and Hidden Constraints

Whenever data is analyzed according to some functional form **DO NOT** calculate the desired result for each data point and then average the individual results! This averaging technique of data analysis has a hidden assumption that the data set includes the origin. Systematic errors or some unexpected additional physical effect may result in a non-zero intercept.

Take for example the theoretical equation

$q = \frac{1}{2} a t^2$ or $t^2 = 1/a\,(2q)$ (data taken as $\theta_i, t_i \pm \sigma_i$)

The best-fit form of this equation should at least include a constant

$t^2 = 1/a(2q) + c$

If a_i is calculated for each data point (θ_i, t_i) by setting $c = 0$; then

$$\frac{1}{\alpha_i} = \frac{t_i^2 - c}{2\theta_i - 0}$$

What is actually being calculated is the slope of a line between the origin and the data point $(2q_i , t_i^2)$.

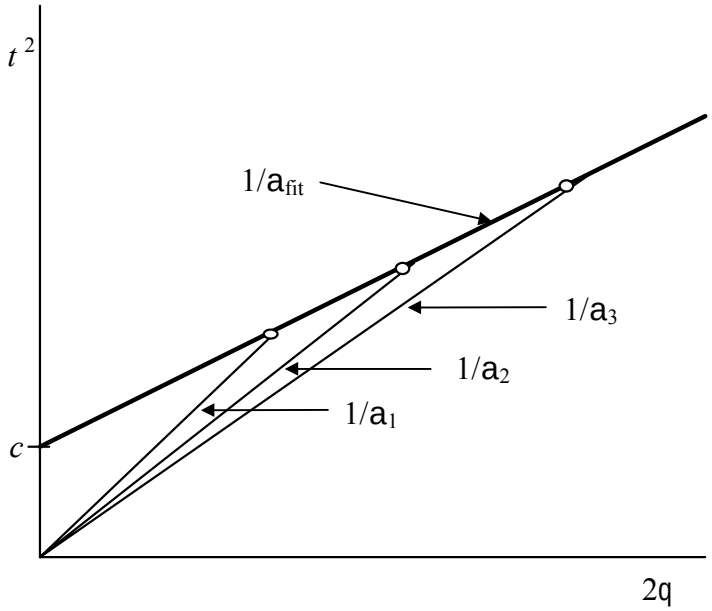
Graphically

It is easily seen that

$$\overline{1/\alpha} = \frac{\sum 1/\alpha_i}{N} \neq 1/\alpha_{\text{fit}}$$

Note the sequence

$1/\alpha_1 > 1/\alpha_2 > 1/\alpha_3$



The best-fit value a_{fit} must come from graphical or least squares analysis. The data should determine the value of the intercept, *c*.

In the case where $\overline{\alpha}$ is calculated with a weighted average the problems in this erroneous technique are further compounded by the additional hidden assumption of a zero sigma on the included (0, 0) point

Even if this averaging procedure happens to produce a value for $\overline{\alpha} = \alpha_{\text{fit}}$ it cannot give a physically meaningful value for s_a. For two parameter data the degree of freedom is *N*-2 even if one of the parameters is assumed to be zero.

Thus

$$\sigma_{1/\alpha} = \frac{1}{N-2}\sum\left(y_i - y_{\text{f}}\right)^2$$

Not

$$\sigma_{\overline{1/\alpha}} = \frac{1}{N-1}\sum\left(1/\alpha_i - \overline{1/\alpha}\right)^2$$

It seems clear that this averaging technique is completely wrong in all aspects and yields no physically meaningful results.

Real physical insight often results from understanding the reasons for an unexpected non-zero intercept. Don't deny yourself this pleasure.

NOT NECESSARY for Ph 3 – VERY NECESSARY for Ph 6&7

The Poisson Distribution

The Poisson distribution represents an approximation to the special case of the binomial distribution, when the average number of events is very much smaller then the total possible number. The number of atoms that spontaneously decay per unit time in a radioactive sample is a classic application of Poisson statistics, as the number of atoms in a sample may be on the order of 10^{20} while the number which disintegrate per unit time interval is many many orders of magnitude smaller.

The uncertainty parameter in Poisson statistics is not due to any measurement error in the time interval or the number of counts, but is a description of the statistical nature of the process observed.

According to Poisson statistics the probability of observing n events per unit time is

$$P(n,\mu) = \frac{\mu^n \mathrm{e}^{-\mu}}{n!} \tag{12}$$

The mean of the distribution μ, is estimated by the familiar

$$\mu \cong \frac{\sum n_i}{N} \qquad (N = \text{number of count intervals}) \tag{13}$$

The standard deviation σ, can be evaluated from the expectation value of the square of the deviations

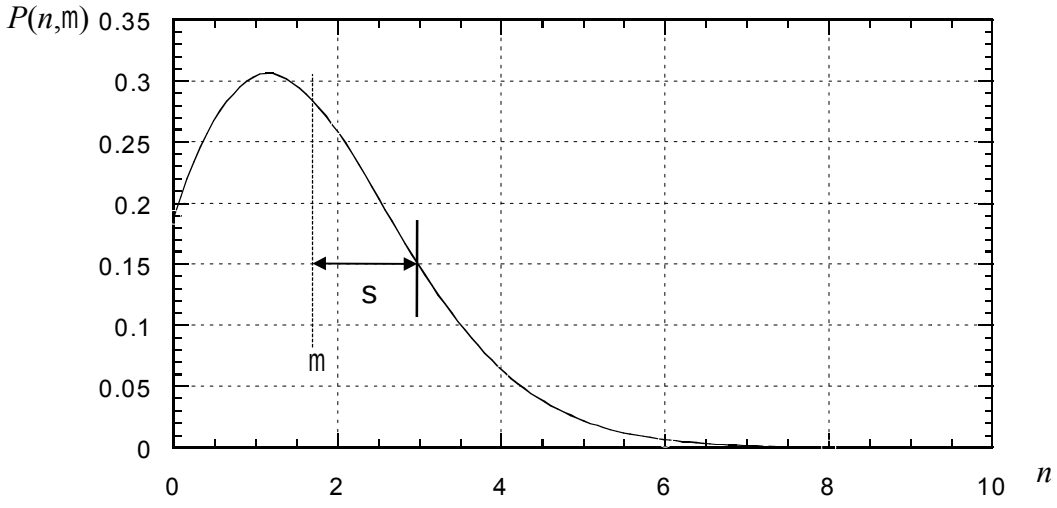
$$\sigma^2 = \langle (n_i - \mu)^2 \rangle = \sum_{n=0}^{\infty} (n_i - \mu)^2 \, P(n,\mu) = \mu \tag{14}$$

using the recurrence formula

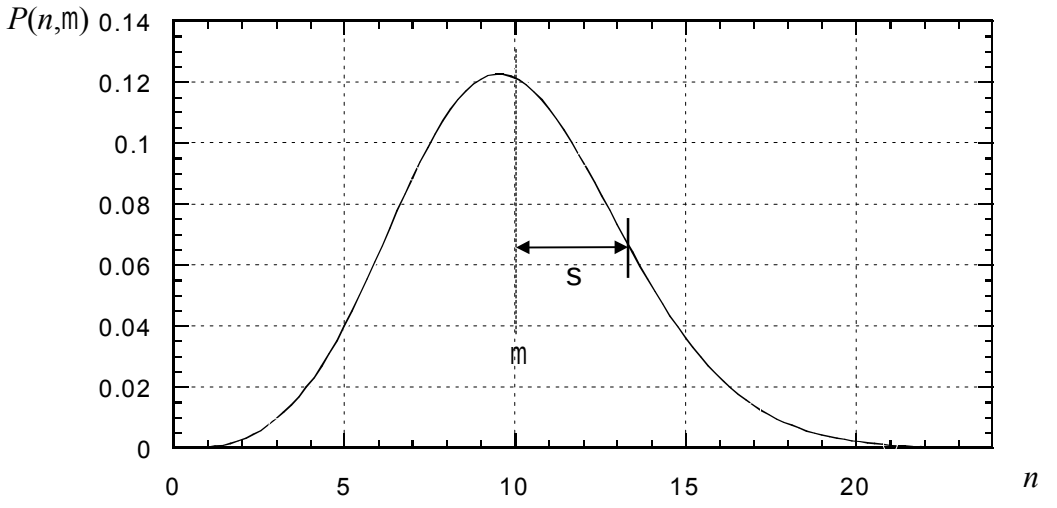
$$P(n+1,\mu) = \frac{\mu}{n+1} P(n,\mu) \qquad \text{where} \quad P(0,\mu) = \mathrm{e}^{-\mu}$$

Thus in a Poisson distribution σ^2 is a fixed parameter in contrast to a Gaussian distribution where it is a free parameter independent of μ. For a discussion of this see Melissinos, Experiments in Modern Physics, page 453.

Poisson Distributions For Two Values Of μ



$m = 1.67$ $s = \sqrt{m} = 1.29$



$m = 10$ $s = \sqrt{m} = 3.16$

Several Comments Regarding the Poisson Distribution

1. For $\mu < 20$ the distribution is asymmetric and the mean, μ, does not coincide with the most probable value for n at the peak of the distribution.

2. For small values of μ the positive and negative values of σ have no simple interpretation as actual deviations. Sigma is an rms deviation, the positive rms deviation being greater than the negative one.

3. For $\mu > 20$ the Poisson distribution becomes symmetric and virtually identical in shape to a Gaussian distribution of the same μ and σ. Also due to the difficulty of calculating the Poisson distribution for large μ the Gaussian distribution equation is usually used.

4. The fact that $\sigma = \sqrt{\mu}$, greatly simplifies the calculation of the standard deviation of the distribution.

The mean, μ, is in general estimated from Equation (13) after N data points are collected. The uncertainty of the mean σ_μ is thus given as

$$\sigma_\mu = \sigma / \sqrt{N} = \sqrt{\mu / N} \tag{15}$$

To show that σ_μ only depends on the <u>total number</u> of counts accumulated, consider two data sets: (a) counts per t sec, (C_i), taken N times

(b) one total count accumulated in T = Nt sec

<u>For Case (a) Using Equations (13), (14), (15)</u>

$$\mu_a = \frac{1}{N}\sum C_i = C_{tot}/N \qquad \text{and} \qquad \sigma_a = \sqrt{\mu_a} = \sqrt{C_{tot}/N}$$

$$\sigma_{\mu_a} = \sigma_a / \sqrt{N} = \sqrt{C_{tot}}/N$$

For the single measurement of Case (b) (T = N*t*)

$$\mu_b \cong C_{\text{tot}} \quad \text{and} \quad \sigma_b = \sqrt{\mu_b} \cong \sqrt{C_{\text{tot}}} \quad \text{giving} \quad \sigma_{\mu_b} = \sigma_b / \sqrt{1} = \sqrt{C_{\text{tot}}}$$

To compare s_m in the two cases, divide m_b and σ_{μ_b} by *N* to get Case (b) results in the same time base as Case (a).

Thus $\mu_b / N = C_{\text{tot}} / N = \mu_a$

and $\sigma_{\mu_b} / N = \sqrt{C_{\text{tot}}} / N = \sigma_{\mu_a}$

Thus the results of Case (b) are the same as Case (a) when stated in the same time base.

Example - Silver Activation Experiment

Background rate recorded in channels 180 to 199

Channel	Counts/(8 sec)	Channel	C_i /(8 sec)	Channel	C_i /(8 sec)
180	52	187	59	194	55
1	46	8	56	5	56
2	61	9	53	6	55
3	60	190	59	7	61
4	48	1	48	8	49
5	55	2	63	199	39
186	53	193	50		

$$\mu \cong \sum C_i / N = 1078/20 = 53.9 \text{ counts / (8 sec)}, \quad \sigma = \sqrt{\mu} = 7.3 / (8\,\text{sec})$$

and $\sigma_\mu = 7.3/\sqrt{20} = 1.6 / (8\,\text{sec})$

Mean background rate *B* = (54 Ò1.6)/(8 sec)

$$\boxed{B = (6.7 \text{ Ò } 0.2)/\text{sec}}$$

Note: The mean rate/sec is NOT $6.7 \pm \sqrt{6.7}$

A Poisson distribution with a mean of 54 and a $\sigma \cong 7$ is symmetric and similar in shape to a Gaussian distribution of the same μ and σ; therefore, approximately 70% of the data should be within Ò1σ of the mean. A scan of the data shows 17 points within that range and 3 points outside that range. The data looks reasonably statistical.

This brings up a point regarding experimental technique. The same statistical uncertainty could have been achieved by simply counting the background once for 160 sec (20 x 8 sec) thereby eliminating the averaging step.

B for 160 sec equals 1072 counts, giving

$$B = (1072 \pm \sqrt{1072}) / (160\,\text{sec})$$

$$B = (1072 \text{ Ò } 32.7) / (160 \text{ sec})$$

$$\boxed{B = (6.7 \text{ Ò } 0.2)/\text{sec}}$$

The limitation of the single long count interval technique is that it <u>assumes</u> the data to be statistical with absolutely no way of checking that assumption. If, during the interval, problems with the data taking equipment arose the result would not necessarily alert you to that fact.

Summary of Equations

Unweighted Case

$$\bar{x} = \frac{1}{N}\sum x_i \cong \mu$$

$$s^2 = \frac{1}{N-1}\sum (x_i - \bar{x})^2 \cong \sigma^2$$

An alternate form of s^2, which is useful for calculators, obtained by expanding $(x_i - \bar{x})^2$, is

$$s^2 = \frac{1}{N-1}\left\{\sum x_i^2 - (1/N)\left(\sum x_i\right)^2\right\}$$

The standard deviation of the mean is $s_{\bar{x}} = s/\sqrt{N}$

A result should be reported as $\boxed{\bar{x} \pm s_{\bar{x}}}$

Weighted Case

$$\bar{x} = \frac{\sum\left(x_i/\sigma_i^2\right)}{\sum\left(1/\sigma_i^2\right)} \qquad \text{and} \qquad \sigma_{\bar{x}}^2 = \frac{1}{\sum\left(1/\sigma_i^2\right)}$$

Forms of the mean and standard deviation least likely to suffer from round off errors on a computer are:

Unweighted

$$M_{i+1} = \left(i\, M_i + x_{i+1}\right)/(i+1) \qquad\qquad s_{i+1}^2 = \frac{(i-1)s_i^2 + i\left(x_{i+1} - M_i\right)^2/(i+1)}{i}$$

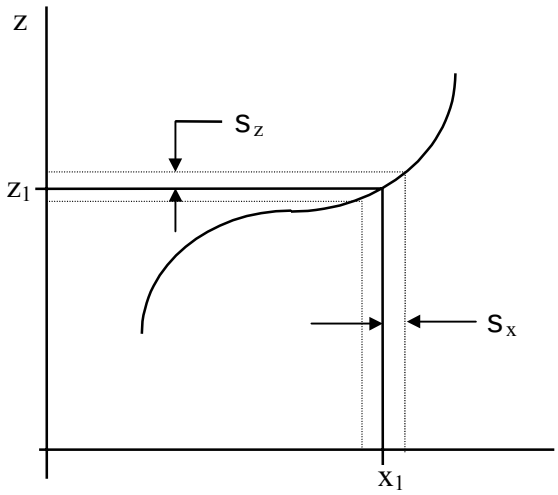
Weighted

$$M_{i+1} = \frac{M_i \sum_i \left(1/\sigma^2\right) + \left(x_{i+1}/\sigma_{i+1}^2\right)}{\sum_{i+1}\left(1/\sigma^2\right)}$$

where $i = \sum_i \left(1/\sigma^2\right)$ and $i+1 = \sum_{i+1}\left(1/\sigma^2\right)$ etc.

Propagation of Uncertainties

It often happens that a quantity in which one is interested is not directly measured, but is computed using other quantities which are directly measured. How does one find the uncertainty in the computed quantity, given the uncertainties in the measured quantities? The situation is illustrated graphically in Figure 1 for the case $z = f(x)$.



Note:

$+s_z$ is not necessarily equal to $-s_z$ even if $+s_x = -s_x$

Figure 1

If the functional form, $f(x)$, is not known, s_z may still be determined by plotting the data and estimating s_z graphically.

A general mathematical technique for evaluating s_z can be suggested by examining a simple case: $z = x^2$

For x between $x_1 + s_x$ and $x_1 - s_x$ the range of z is $(z_1 + s_z)$ and $(z_1 - s_z)$

Thus $z_1 \pm s_z = (x_1 \pm s_x)^2$

Giving $z_1 \pm s_z = x_1^2 \pm 2x_1 s_x + s_x^2 = x_1^2 [1 \pm 2(s_x/x_1) + (s_x/x_1)^2]$

If s_x/x_1 is small ($<<1$) then $(s_x/x_1)^2$ can be ignored.

Thus $s_z = 2x_1 s_x$

This suggests that in general

$$\sigma_z = \frac{df(x)}{dx}\sigma_x \tag{1}$$

Equation (1) is in fact correct provided the slope is reasonably constant over the distance s_x. For the more general case where $z = f(x, y, \ldots)$, we have

$$\Delta z = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \ldots$$

To avoid the possibility of accidental cancellations as well as for more sophisticated mathematical reasons, the standard practice is to calculate the average of the square of Δz.

$$\overline{(\Delta z)^2} = \left(\frac{\partial f}{\partial x}\right)^2 \overline{(\Delta x)^2} + \left(\frac{\partial f}{\partial y}\right)^2 \overline{(\Delta y)^2} + 2\left|\frac{\partial f}{\partial x}\right|\left|\frac{\partial f}{\partial y}\right| \overline{\Delta x \Delta y} + \ldots \tag{2}$$

Again the partial derivatives are assumed to be approximately constant over the range of the averaging. Note the absolute values in the third term.

The third term on the right-hand side of equation (2) is the covariant term; its average evaluates any interdependence between x and y. If the random variations in x and y are uncorrelated, then averaging the product $\Delta x\, \Delta y$ should include as many negative contributions as positive, and the average will therefore vanish.

Thus we are left with

$$\sigma_z^2 = \overline{(\Delta z)^2} = \left(\frac{\partial f}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial f}{\partial y}\sigma_y\right)^2 + \ldots \tag{3}$$

Rewriting Equation (3) in summation notation gives

$$\boxed{\sigma_z^2 = \sum_{i=1}^{n}\left(\frac{\partial f}{\partial v_i}\right)^2 \sigma_{v_i}^2} \qquad (n = \#\text{ of variables}) \tag{4}$$

where $z = f(v_i) = f(v_1, v_2, \ldots, v_n)$.

As a check of these ideas evaluate s_z in terms of s_x for $z = x^3$ using Equation (1); and then do likewise for $z = xy$ with $y = x^2$ using Equation (3). Why is there a difference between the two results?

NOTE: $$\overline{\Delta x \Delta y} = \sigma_{xy} = \frac{1}{N-1}\Sigma(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N-1}\left[\Sigma(x_i - y_i) - N\bar{x}\bar{y}\right]$$

Table 1

Formulas for the Propagation of Uncertainties

Functional Form	Error Formula
y = ax Ò bz	
y = Ò axz	
y = Ò ax/z	
y = Ò $ax^{Ò b}$	
y = a ln(Òbx)	
y = f(x, z)	

Assumes $y = f(x, z)$, with x and z **uncorrelated**

The partial differential symbols, ×*f*/×*x* etc., on the previous page are executed by differentiating $f(x, y, z)$ with respect to one variable, say x, while holding the other variables, y & z, constant.

To demonstrate this let's calculate s_y for $y = axz$, in Table 1.

$$\sigma_y^2 = \left(\frac{\partial y}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial y}{\partial z}\sigma_z\right)^2$$

Now the ×y/×x holding z constant is $\frac{\partial y}{\partial x} = az$, and likewise $\frac{\partial y}{\partial z} = ax$

Thus $\sigma_y^2 = (az)^2\sigma_x^2 + (ax)^2\sigma_z^2$

or in fractional form $\boxed{\frac{\sigma_y^2}{y^2} = \frac{\sigma_x^2}{x^2} + \frac{\sigma_z^2}{z^2}}$

Example 1 - Standard Deviation of x^2

A certain length is measured and is found to be x = 6.0 Ò 0.1 cm. What is the uncertainty in x^2? (i.e., how much uncertainty is there in the area of a square when the area is calculated from a measurement of one side ?)

CASE -1 Let $A = x^2$

Then $\frac{dA}{dx} = 2x = 2(6.0) = 12.0\,\text{cm}$

Therefore, $\sigma_A = 2x\sigma_x$ = 2(6.0)(0.1) = 1.2 cm^2.

Thus one would write A = 36 Ò 1.2 cm^2 for the area of an assumed square 6.0 Ò 0.1 cm on a side.

CASE -2 If adjacent sides of the square are measured, is the uncertainty in the area changed?

Here $A = xy$.

$\sigma_A^2 = y^2\sigma_x^2 + x^2\sigma_y^2$

If $x = y$ and $\sigma_x = \sigma_y$

$\sigma_A = \sqrt{2}\,x\sigma_x$

This is a smaller uncertainty than that calculated for CASE -1. Some precision has been gained by measuring the second side rather than assuming the adjacent sides equal.

CASE -3 The same reduced uncertainty could be obtained by twice measuring only one side and again assuming the adjacent side equal.

$\bar{x} = \sum x_i / N \quad (N = 2) \qquad and \quad \sigma_{\bar{x}} = \sigma_x / \sqrt{2} \qquad (\text{where } \sigma_1 = \sigma_2 = \sigma_x)$

$A = \bar{x}^2 \qquad \text{Giving} \quad \sigma_A = 2\bar{x}\sigma_{\bar{x}} = 2\bar{x}\,\sigma_x / \sqrt{2} \quad \text{or} \quad \sigma_A = \sqrt{2}\,\bar{x}\sigma_x$

Thus the mathematics is simply giving credit for making more than one measurement, but is unable to tell which technique is better. THAT IS THE EXPERIMENTER'S RESPONSIBILITY. Clearly a technique that only measures one side of a square and assumes the other side equal is less reliable than a technique which assumes nothing and measures the adjacent sides.

Example 2 - Detector Dead Time

The dead time of a radiation detector can be obtained from the two source technique. The counting rates, R_1 and R_2, of the individual sources is compared to the combined rate, R_{12} , of the two sources together. The dead time should show up as a difference between the sum of the individual rates, $R_1 + R_2$, and the combined rate R_{12}.

The equation that relates these quantities to the dead time, t_d , is

$$t_d \cong \frac{R_1 + R_2 - R_{12}}{2R_1R_2}$$

Twenty measurements of the quantities R_1, R_2, and R_{12} were made and listed as $\overline{R} \pm \sigma/\sqrt{N}$ $\left(\sigma = \sqrt{\overline{R}}\right.$ for a statistical process like radioactive decay$\left.\right)$ [See page 1-11, Eq. (4)]

$R_1 = 10206$ Ò 22 counts/sec $R_2 = 8340$ Ò 20 c/sec $R_{12} = 18258$ Ò 30 c/sec

Rather than methodically applying Equation (3) to obtain the uncertainty in t_d , I am going to make a simplifying assumption and calculate the uncertainty of the numerator and denominator separately. The result will tell me if this is a reasonable approach.

$$t_d \cong \frac{(10206 + 8340 - 18258) \pm \sqrt{(22)^2 + (20)^2 + (30)^2}}{2(10206)(8340) \pm \sqrt{2^2\left[R_1^2\left(\sigma_{R_1}\right)^2 + R_2^2\left(\sigma_{R_2}\right)^2\right]}}$$

$$t_d \cong \frac{288 \pm 42}{(1.702 \pm 0.005) \times 10^8}$$

It is clear that we know the denominator much more accurately than we know the numerator and may therefore consider the denominator a constant, justifying the simplifying assumption. The uncertainty in t_d should therefore be the same percentage as the uncertainty in the numerator (i.e. 14%) giving $t_d \cong 1.69\text{x}10^{-6}$ Ò 14% or

$$\boxed{t_d \cong (1.69 \text{Ò} 0.23) \text{Õ} 10^{-6} \text{ sec}}$$

NOTE: This is an excellent example of the dangers involved in results that depend on differences. The uncertainties in R_1 and R_2 are at worst Ò0.24%, but $\sigma_{t_d} = \pm 14\%$.

Example 3a - Transmission Cable

The note next to Figure 1, page 2-1, suggests that the mathematical formalism of Equation (1), page 2-2, can break down. If the second derivative of a function *f(x)* is large in the region of interest,the magnitudes of $+\sigma_x$ and $-\sigma_x$ may not be equal.

Take the relation $V = V_o\, e^{-x/x_a}$ for the voltage at x in a transmission cable. Then if we measure V_o and V at $x = L$ we can calculate the attenuation length x_a from

$$x_a = \frac{-L}{\ell n\left(V/V_o\right)} \tag{5}$$

Applying Equation (1) gives $\sigma_{x_a} = \dfrac{L}{\left(\ell n V/V_o\right)^2} \times \dfrac{V}{V_o} \times \dfrac{\sigma_V}{V_o}$

The result of a typical measurement might be $V/V_o = 0.95 \pm 0.04$ giving

$$x_a / L = 19 \pm 14.$$

By direct substitution into Equation (5) we get the correct result

$$x_a / L = 19^{+80}_{-9} \qquad \begin{pmatrix} \text{upper limit for } V/V_o = 0.99 \\ \text{lower limit for } V/V_o = 0.91 \end{pmatrix}$$

A plot of the function clearly shows what is happening.

3b.

The direct substitution idea is quite helpful at times when the uncertainty cannot easily be expressed in analytic form. For instance

$$\theta = \sin^{-1}\left(k\lambda / d\right) = \sin^{-1} A \qquad (\lambda = 3.2 \pm 0.2 \text{ cm},\ d = 10.2 \pm 0.1 \text{ cm})$$

Rather than having to remember the differential of the inverse sine, just apply Equation (3) to the argument, $A = k\lambda/d$, and get σ_θ by direct substitution.

$$\frac{\sigma_A^2}{A^2} = \frac{\sigma_\lambda^2}{\lambda^2} + \frac{\sigma_d^2}{d^2} \qquad (k = 1)$$

$$= 3.9\times10^{-3} + 9.6\times10^{-5}$$

$$= 3.99\times10^{-3}$$

Now $A = (1)(3.2)/(10.2) = 0.31$ giving $\sigma_A = 0.02$

$$\theta = \sin^{-1}(0.31 \pm 0.02) = \begin{cases} 19.5^\circ & +\sigma_\theta = 19.5 - 18.3 = 1.2^\circ \\ 18.3^\circ & \\ 17.1^\circ & -\sigma_\theta = 18.3 - 17.1 = 1.2^\circ \end{cases}$$

Thus $\theta = 18.3^\circ \pm 1.2^\circ$

Example 4 - Relative Uncertainties $(\sigma_f / f)^2$

In designing an experiment in which the result is the consequence of several measurements, assessing the impact of the various measurements on the result is often useful in defining the weakest link in the measurement chain. Take the case of the Kelvin Absolute Voltmeter equation $V^2 = km$ with

$$k = \frac{2d^2 g}{\varepsilon_0 \pi r^2} \quad \text{and} \quad \begin{aligned} d &= 0.309 \pm 0.001\,\text{cm (plate spacing)} \\ r &= 3.073 \pm 0.001\,\text{cm (plate radius)} \end{aligned}$$

Applying Equation (3) yields

$$\left(\frac{\sigma_k}{k}\right)^2 = \left(\frac{2}{d}\right)^2 \sigma_d^2 + \left(\frac{2}{r}\right)^2 \sigma_r^2$$

$$= 4.18 \times 10^{-5} + 4.23 \times 10^{-7} = 4.22 \times 10^{-5}$$

$$\frac{\sigma_k}{k} = 0.65 \times 10^{-2} \qquad \text{or} \qquad 0.65\%$$

How accurately do the constants need to be known? Is g = 980 cm/sec^2 adequate? (g = 979.577 ± 0.002 at Caltech). Trying g = 979.5 ± 0.1 cm/sec^2 gives

$$\frac{\sigma_k}{k} = \left(\frac{1}{k}\frac{\partial k}{\partial g}\right)\sigma_g = \frac{1}{g}\sigma_g = \left(\frac{1}{979.5}\right)(0.1) = 1.0 \times 10^{-4}$$

This is negligible compared to the contribution from the plate spacing, d. Thus it is clear that if we wanted to know k to better than 0.65%, we would have to make a more accurate measurement of d.

Example 5 - Counting Rate

The rate *R*, from a radioactive source plus background was measured four times giving 848, 815, 833, and 825 counts/min. The background rate *B*, measured four times gave 365, 372, 392, and 358 counts/min. What was the rate and its uncertainty from the source alone?

$$\overline{R} = 830 \pm \sqrt{830}/\sqrt{4} = 830 \pm 14$$
$$\overline{B} = 371 \pm \sqrt{371}/\sqrt{4} = 371 \pm 9$$

[See page 1-11, Eq. (4)]

giving $S = \overline{R} - \overline{B} = 459$ and $\sigma = \sqrt{(14)^2 + (9)^2}$

Thus

$S = 459 \pm 17$ counts/min

Example 6 - Comparing Results

To compare two determinations of the same quantity calculate the difference and its uncertainty.

$Q_1 = 1.79 \pm 0.03$ $Q_2 = 1.68 \pm 0.06$

$$\Delta = |Q_1 - Q_2| \pm \sigma_\Delta = |1.79 - 1.68| \pm \sqrt{(0.03)^2 + (0.06)^2}$$

giving

$D = 0.11 \pm 0.067$

Note that D is within 2s of zero, suggesting that these <u>are</u> two determinations of the same quantity.

NOTE: This technique will be required whenever multiple values are compared

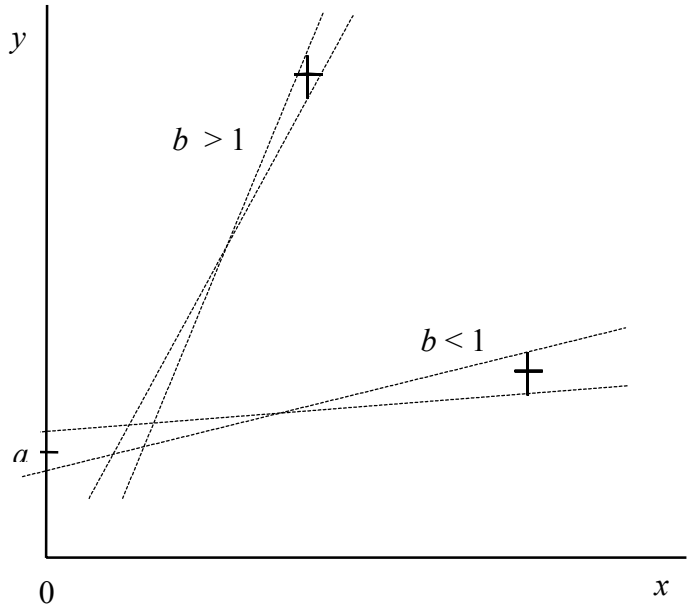
The Impact of Uncertainties, σ_y & σ_x

The discussion so far has focussed on the uncertainties associated with a measured quantity and it's impact on the results calculated from the measurements. The more common measurements are of pairs of quantities where the uncertainties on each measured quantity can be assessed. Such data may fit a form such as $y = a + bx$, $y = ae^{bx}$ etc. Let's look at the sensitivity of a particular fit to uncertainties in x relative to the usual uncertainties in y.

Let's start with the linear case $\boxed{y = a + bx}$. Solving for x and differentiating gives

$$\sigma_x = \frac{\sigma_y}{b} . \qquad (6)$$

This implies that if b is large the equivalent s_x, which has a similar effect on the fit as s_y, is some fraction of s_y. If b is small, but greater than one, the equivalent s_x is similar to s_y. If b is less than one the fit is dominated by s_y, becoming less and less sensitive to s_x as b approaches zero. See the figure below.



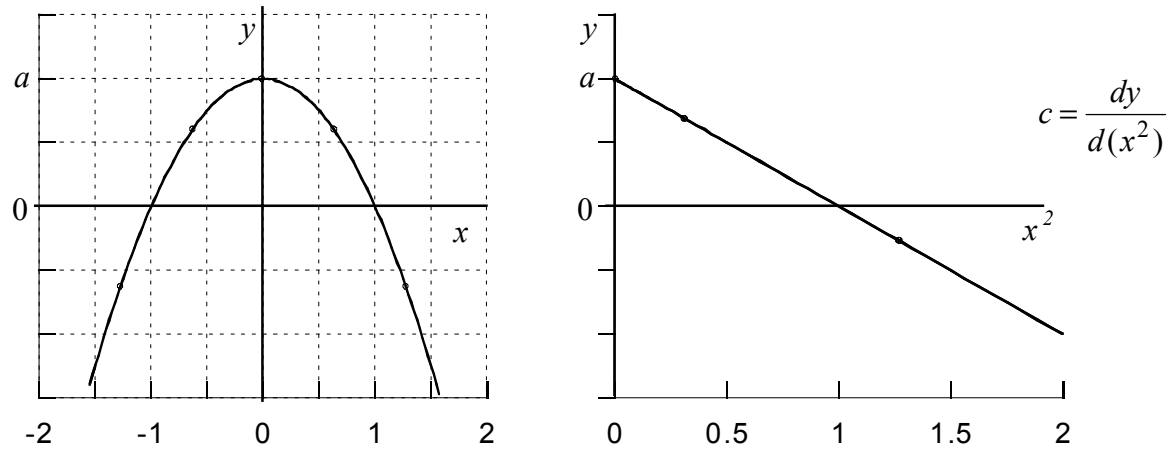
Next let's look at exponential data described by $\boxed{y = a\,e^{bx}}$. Taking the log, solving for the independent variable x, and differentiating, gives $\sigma_x = \frac{1}{b}\frac{\sigma_y}{y}$. (7)

The conclusions regarding the effectiveness of s_x, for various values of b, are similar to the conclusions in the linear case, with the additional sensitivity to s_x due to the $1/y$ factor. Thus with exponential data the uncertainties s_a and s_b in the fit parameters a and b may be dominated by the sensitivity to s_x. **Uncertainties in the Independent variable are always IMPORTANT in explaining the scatter in the data and in determining reasonable values for the uncertainties in the fit parameters such as σ_a and σ_b.**

Graphing

A plot of experimental data is always an appropriate first step in data analysis. The plot will show trends in the data, how well the individual points fit the trend, and if the data set has gaps in it that would be desirable to fill. In general a plot of your data will convey a clearer picture of your results than just a table of numbers.

Plots of the data which exhibit a linear relationship, ($y = a + bx$), between variables are particularly useful. Take for instance measured data points x_i , y_i, where there may be a quadratic relationship between x and y (i.e. $y = a + cx^2$). A plot of y_i versus x_i would show the parabolic shape of the relationship between x and y but would not be useful in determining the constants a and c. A plot of y_i versus x_i^2 would allow a linear fit, averaged over all the data, resulting in a best-fit value for a from the y axis intercept, and c from the slope of the fit.



As another example of the advantages gained by choosing coordinate variables that linearize

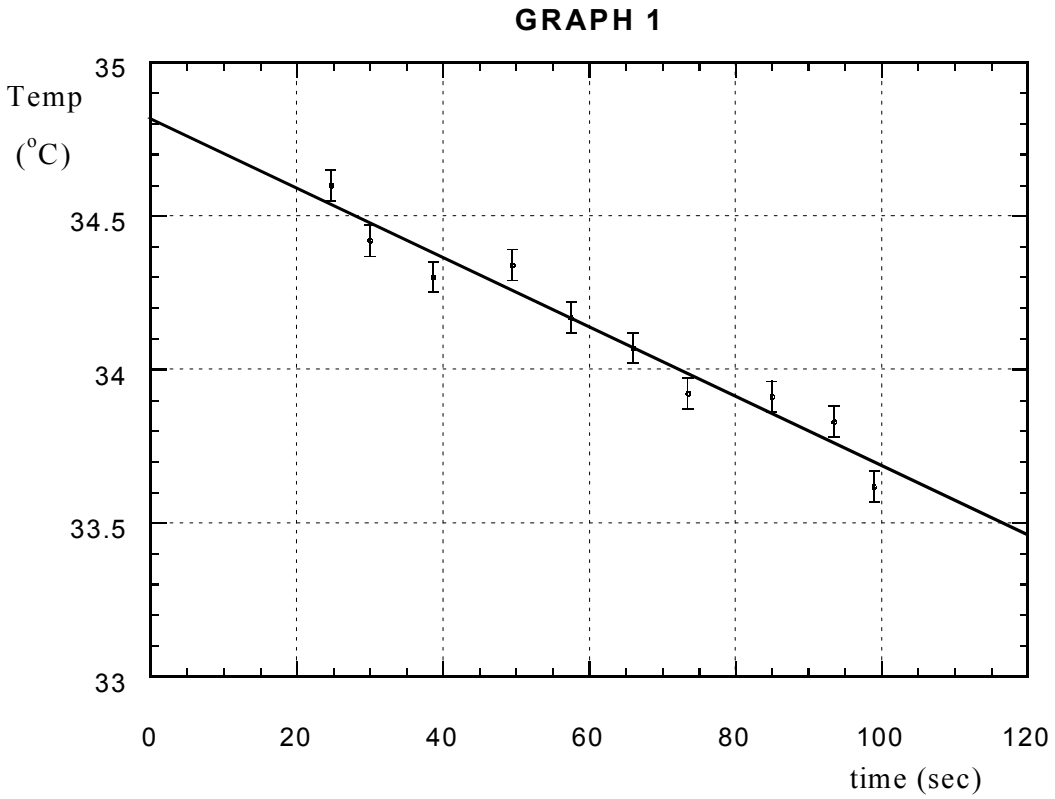
a functional form, take Sellmeier's equation relating index of refraction to wavelength.

$$n^2 = 1 + A\lambda^2 / \left(\lambda^2 - B\right)$$

Rewritten in linearized form this equation becomes

$$\lambda^2 = B + A\left(\lambda^2 / (n^2 - 1)\right)$$

A plot of λ^2 versus $\lambda^2 /(n^2 - 1)$ will yield best-fit values for A and B from the slope and intercept.



Graph 1 is a linear plot of some typical data. Several things about Graph 1 should be noted: ($T = a + bt$)

1. The axes are clearly labeled as to quantity measured and it's units.
2. The scale of the axes was chosen so that the plotted data extends over a reasonable fraction of the page. (Do not cramp all your data into one corner of a plot.)
3 The uncertainties on the individual temperature measurements have been displayed as error bars, which display the two limits $T_i + s_i$ and $T_i - s_i$.
4 The best-fit line was drawn through the data, in such a way as to give an equal and minimum dispersion of points on either side of the line. A clear plastic ruler is very helpful for this. Further note that you need not feel constrained to make the best-fit line pass through every experimental point. If your uncertainty estimates are valid, about 1/3 of the error bars <u>should miss</u> the best-fit line.

Graphical Analysis

Data plotted in Graph 1.

Time(sec)	Temp(^{o}C)
24.7 Ò 0.2	34.60 Ò 0.05
30.0	34.42
38.7	34.30
49.5	34.34
57.5	34.17
66.0	34.07
73.5	33.92
85.0	33.91
93.5	33.83
99.0	33.62

From the appearance of the plotted data in Graph 1 a linear relationship between time and temperature seems reasonable. The slope of the best-fit line will give the time averaged cooling rate. To find the slope we calculate Dy/Dx. This should be done over the full range of the plotted data to minimize the uncertainty in the calculated slope due to reading errors at the chosen points. (best-fit line T = a + $b$$t$)

$$b = \frac{\Delta T}{\Delta t} = \frac{T_1 - T_2}{t_1 - t_2} = \frac{34.70 - 33.45}{10.0 - 120.0} = -0.0113\ ^{\circ}C / \sec$$

The sigma on the slope can be estimated by ignoring the uncertainty in the time and using just the temperature uncertainty s_T = Ò 0.05 ^{o}C

$$s_b = \frac{\sqrt{(0.05)^2 + (0.05)^2}}{110\ \sec} = 0.0006\ ^{o}\text{C/sec}$$

NOTE: This estimate is only reasonable if the uncertainties in the temperature are consistent with the last sentence of point 4, under Graph 1. An unweighted LSF to the above data suggests that an uncertainty in the temperature of 0.07 ^{o}C would be more consistent with the scatter in the data. The same LSF analysis shows the sigma on the slope equal to 0.0008 ^{o}C/sec.

A Final Point: In both plotting and analyzing the data the sigma on the time has been ignored. The reason is clearly seen in Graph 1. The sigma on the time is insignificant over the range of the time data (24 to 100 sec) compared to the sigma on the temperature over it's range (33.6 to 34.6 ^{0}C)

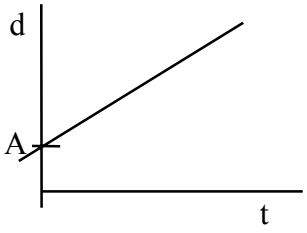
i.e. (0.2 / 76) x 100 = 0.26% whereas (0.05 / 1) x 100 = 5%

Thus the uncertainty in the results will be dominated by the uncertainty in the temperature.

NOTE

Relating the constants (A & B) of the fit equation y = A + B*x* to the constants in a theoretical equation.

Example: Theoretical equation $d = d_o + v_o t$

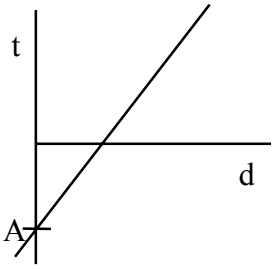
If the data is plotted as 

and fitted with y = A + Bx

Figure 1

with y = d and x = t, then clearly the y axis intercept $A = d_o$ and the slope $B = v_o$.

Whereas:

If the data is plotted as 

and again fitted with y = A + Bx

Figure 2

then the theoretical equation must be rewritten as $t = 1/v_o(d - d_o)$ or $t = -(d_o/v_o) + (1/v_o)d$ in order to relate A & B to d_o and v_o. For this plot of the variables clearly the *y* axis intercept $A = -(d_o/v_o)$ and the slope $B = (1/v_o)$.

Why is this additional complication necessary? Why not just plot the data in the most straightforward manner? The reason is related to how most fitting programs are written on computers; namely that the *y* variable plotted on the vertical axis MUST be the variable with uncertainties. (See figure page 4-1)

Thus, in the above example if the data is in the form d , t Ò s_t you must plot the data in the form of Figure 2 in order to obtain a weighted fit of the form y = A + Bx.

Difference Plots

Many times the range of a measured variable is so much greater than the measurement uncertainties that the error bars are invisible on the plot of the data. In such a case the precision of the data is not visible, and any visual assessment of the goodness of fit relative to the best-fit line is lost. In such cases a difference plot is essential. In a difference plot the large variation in a measured quantity is removed by plotting the difference between the measured quantities and the corresponding point on the best-fit line. On this scale the error bars will now be visible and a goodness of fit assessment, as well as an estimate of the uncertainties on the slope and intercept, may be made. An excellent example of these ideas is provided by the precession data from the "Maxwell Top Experiment".

Sample Data - Plotted in Graph 2

d (cm)	Ω (radian/sec)
8 Ò 0.02	0.13900 Ò 0.00007
9	.12419 .00005
10	.10743 .00003
12	.07671 .00001
14	.045004 .000002
19	-0.033314 .000001
21	-0.064355 .000006

In order to determine the significant uncertainty calculate

$$\frac{\sigma_d}{\text{Range }(d)} \times 100 = \frac{0.02}{13} \times 100 = 0.15\%$$

and

$$\frac{\sigma_\Omega}{\text{Range }(\Omega)} \times 100 = \frac{0.00002}{0.20330} \times 100 = 0.01\%$$

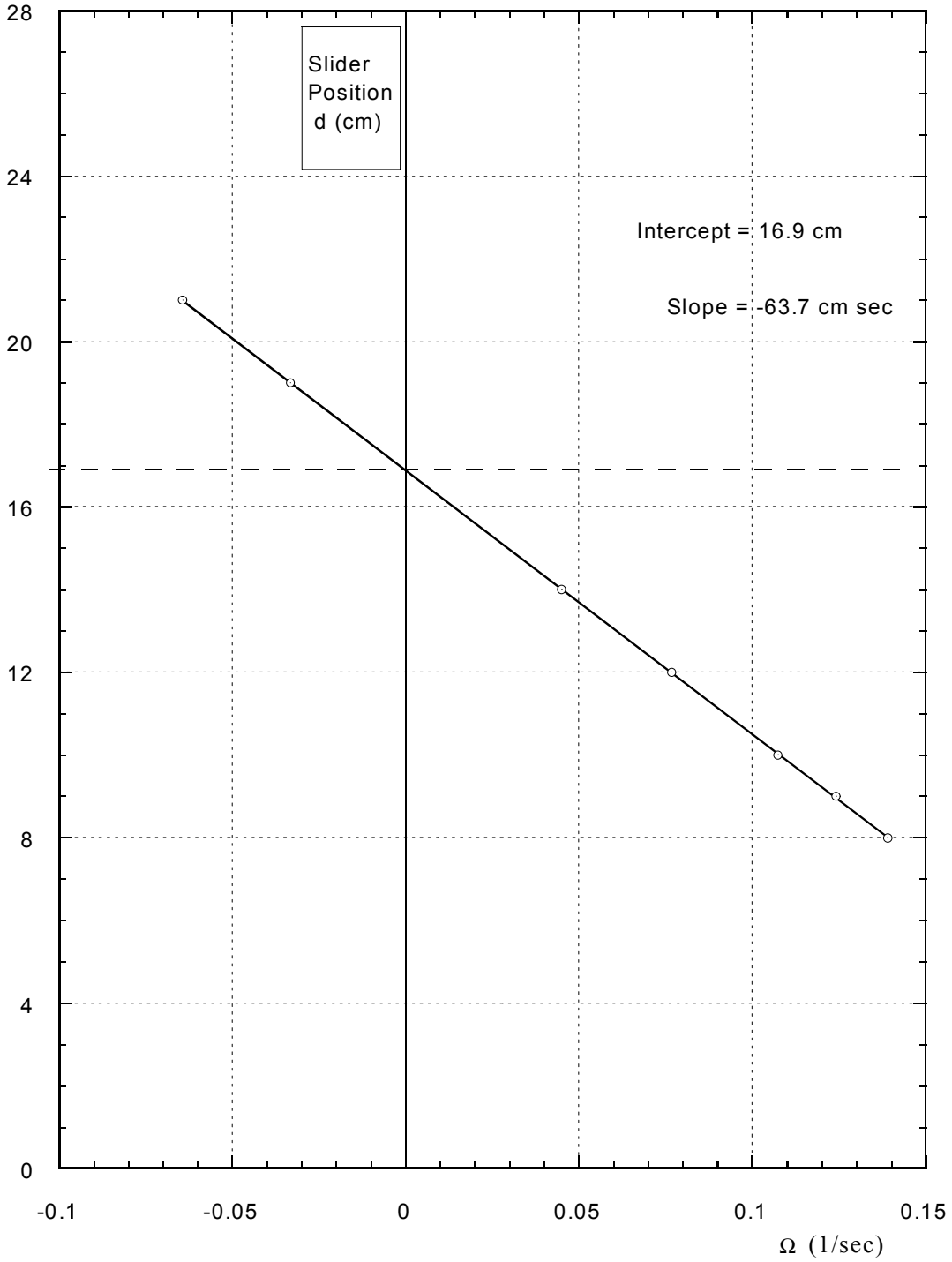
Thus the plot in Graph 2 is W v.s. d Ò s_d.

From Graph 2 the 1st order fit equation is $y_f = A + B\,W$

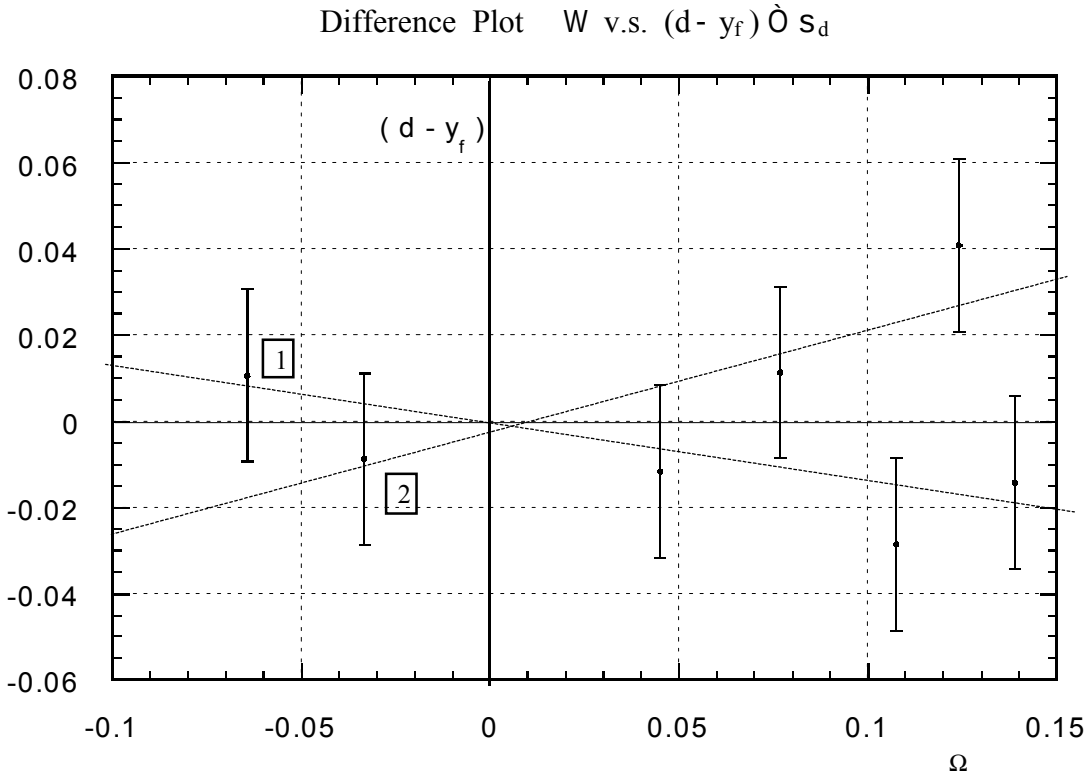
$$y_f = 16.9 - 63.7\,W = A + B\,W$$

The Difference Plot (Graph 3) is a plot of W v.s. (d - y_f) Ò s_d. A typical data point is W = 0.139/sec, d - y_f = - 0.045 Ò 0.02cm.

GRAPH 2



GRAPH 3



The dashed lines on Graph 3 represents an estimate of the max. and the min. permissible slope, from which an estimate of the uncertainties on A and B may be made. The estimates of the sigmas are going to vary depending on what each individual decides is a max. and min. permissible slope, but in general will be too large by at least a factor of two. These estimates are therefore not for quoting but are merely an "in house" 1st order attempt. A quotable result would have to come from LSF analysis (see Chapter 4).

From Graph 3 taking the average difference of the intercepts (max. & min.)

$$\sigma_A = \tfrac{1}{2}\left|0.0-(-.0025)\right| = 0.001$$

$$\sigma_B = \tfrac{1}{2}\left|\text{max. slope} - \text{min. slope}\right|$$

$$= \tfrac{1}{2}\left|0.19-(-0.13)\right| = 0.16$$

Thus A = 16.90 Ò 0.001 cm, B = - 63.80 Ò 0.16 cm-sec

The difference plot has thus provided several things. The estimates of the uncertainties on A and B have been obtained and equally as important the difference plot has given a visual assessment of the goodness of fit.

A Final Comment

The difference plot displays the residual information in the data after the initial fit form information has been extracted. When all the information has been extracted from a data set the points on the difference plot should be randomly scattered about the x axis of the difference plot.

UNITS - In any fit equation the question of the units of the fit parameters may arise. Take for example $Y = A + Bx + C\,x^2$. The equal sign must apply to the units as well as the numerical value of each side. If the units of Y are meters then the units of $C\,x^2$ must also be meters. If the units of x are seconds this dictates that the units of C must be m/sec^2.

χ^2 Goodness of Fit Test

Analogous to Equation (7) page 1-4 define Chi^2 as

$$\chi^2 = \sum \frac{\left(y_i - y_f\right)^2}{\sigma_i^2}$$

which is a goodness of fit test for a line drawn through a set of data. A more generally useful quantity, which does not require a table of probabilities to interpret it, is a quantity, normalized over the number of data points *N*, called the Reduced Chi^2 defined as

$$\boxed{\frac{\chi^2}{N-2} = \frac{1}{N-2}\sum \frac{\left(y_i - y_f\right)^2}{\sigma_i^2}}$$

From Graph 3 page 3-7, point 1 is approximately 1/2 sigma unit (a sigma unit being defined by the size of the error bar for that point) from the best-fit line, and point 2 is aproximately 1/2 s from the best-fit line. The contribution of the rest of the 7 points to $\chi^2/N-2$ can be similarly assessed, giving

$$\frac{\chi^2}{N-2} = \frac{1}{5}\left[(0.5\sigma_1)^2/\sigma_1^2 + (0.5\sigma_2)^2/\sigma_2^2 + (0.5\sigma_3)^2/\sigma_3^2 + \ldots\right] = 1.5$$

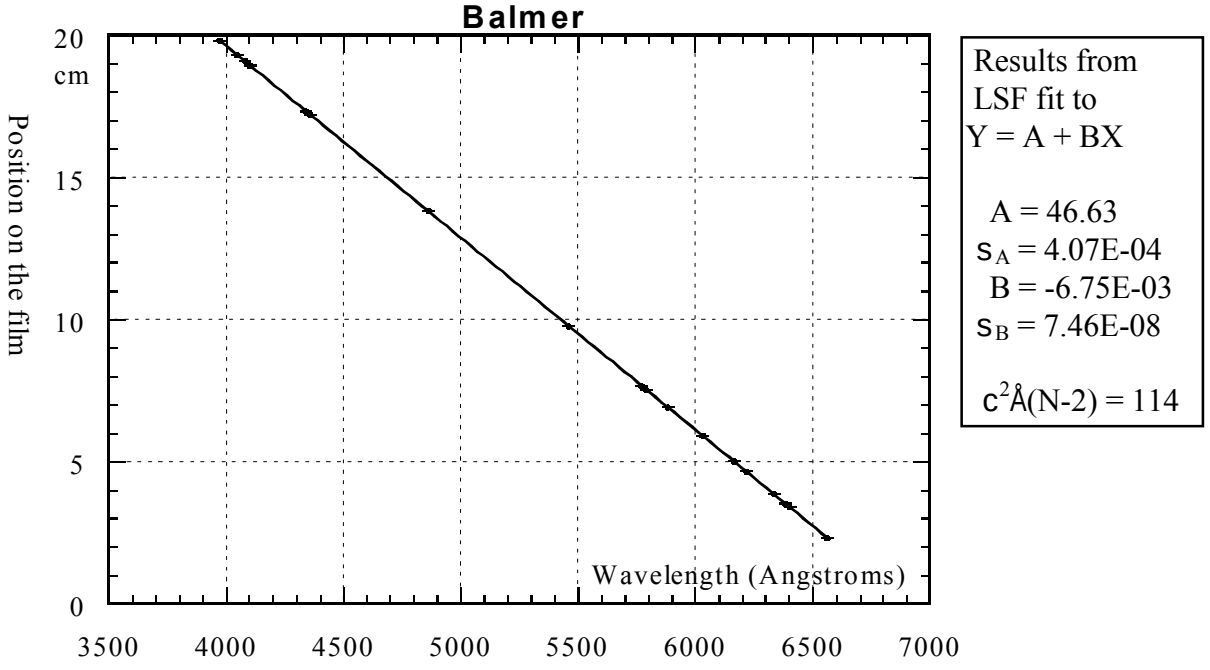
For an ideal fit, one in which the data is randomly scattered about the best-fit line and the size of the error bars are consistent with the scatter, the value of $c^2/N-2$ is 1. For $N < 20$ the ideal value is slightly less than one.

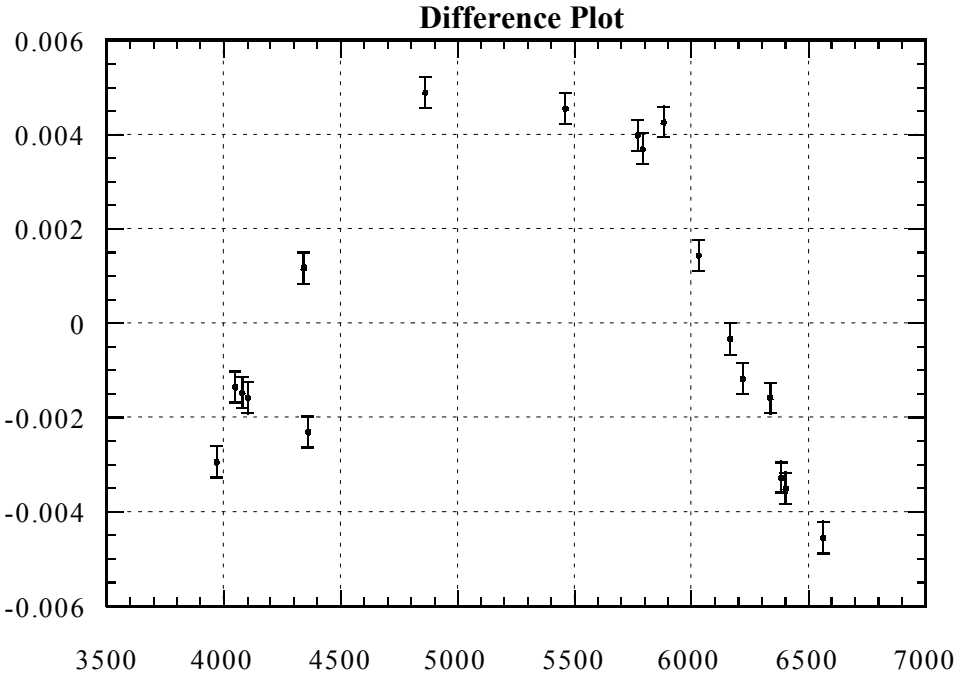
Possible reasons for a Reduced χ^2 not being 1:

a) Error bars too small for the scatter in the data giving a Reduced Chi$^2 > 1$.
b) Error bars too large relative to the scatter in the data giving a Reduced Chi$^2 < 1$
c) A bad point which may make a large contribution to the Reduced Chi^2.
d) Second order effects that leave residual information in the data set after the parameters of the primary effect have been extracted, leading to a $c^2/N-2 > 1$ (See page 3-10 for a prime example of d).

In any case if the value of $c^2/N-2$ is much greater than, or much less than one, a closer scrutiny of the data is indicated. Finally the method used for fitting the data was developed from Gaussian assumptions about the data; the c^2 test closes the loop by testing the data and the results to see if they meet the original Gaussian criteria.

Another example of the usefulness of difference plots is demonstrated by the spectrographic data plotted below. The presence of the second order cosine effect in the data is suggested in the first plot by the rather large value of $c^2/(N\text{-}2)$; but is clearly shown in the difference plot.





Logarithmic Graph Paper

Physical quantities encountered in the lab may be related by some type of power law: e.g. $y = bx^c$. A graph paper that would allow us to determine b and c from the experimental data would be useful. Consider what would happen if we take the logarithm of both sides of $y = bx^c$.

$$\ln(y) = \ln(bx^c) \quad \text{or} \quad \ln(y) = \ln(b) + c\ln(x)$$

This is now of the form, $w = a + cz$. If we plot this on paper that has both axes ruled logarithmically, we should get a straight line with a slope c. As there is no origin (point 0,0) on log-log paper the value of b must be found by some method other than the y axis intercept at $x = 0$; for instance at the intercept of the $x = 1$ line, if it exists [$\ln(1) = 0$ thus $\ln(y) = \ln(b)$ or $b = y$]. For example, consider the following set of experimental data:

LogLog

x	y
1.2 Ò 0.01	3.0 Ò 0.3
1.9 Ò 0.02	6.0 0.6
3.0 0.03	14.1 1.4
4.6 0.05	36.2 3.6
8.1 0.08	85.2 8.5

Graph 4 is a plot of the data on log-log paper.

The points are fitted by the straight line shown in Graph 4 which demonstrates that y does indeed obey a relationship of the form

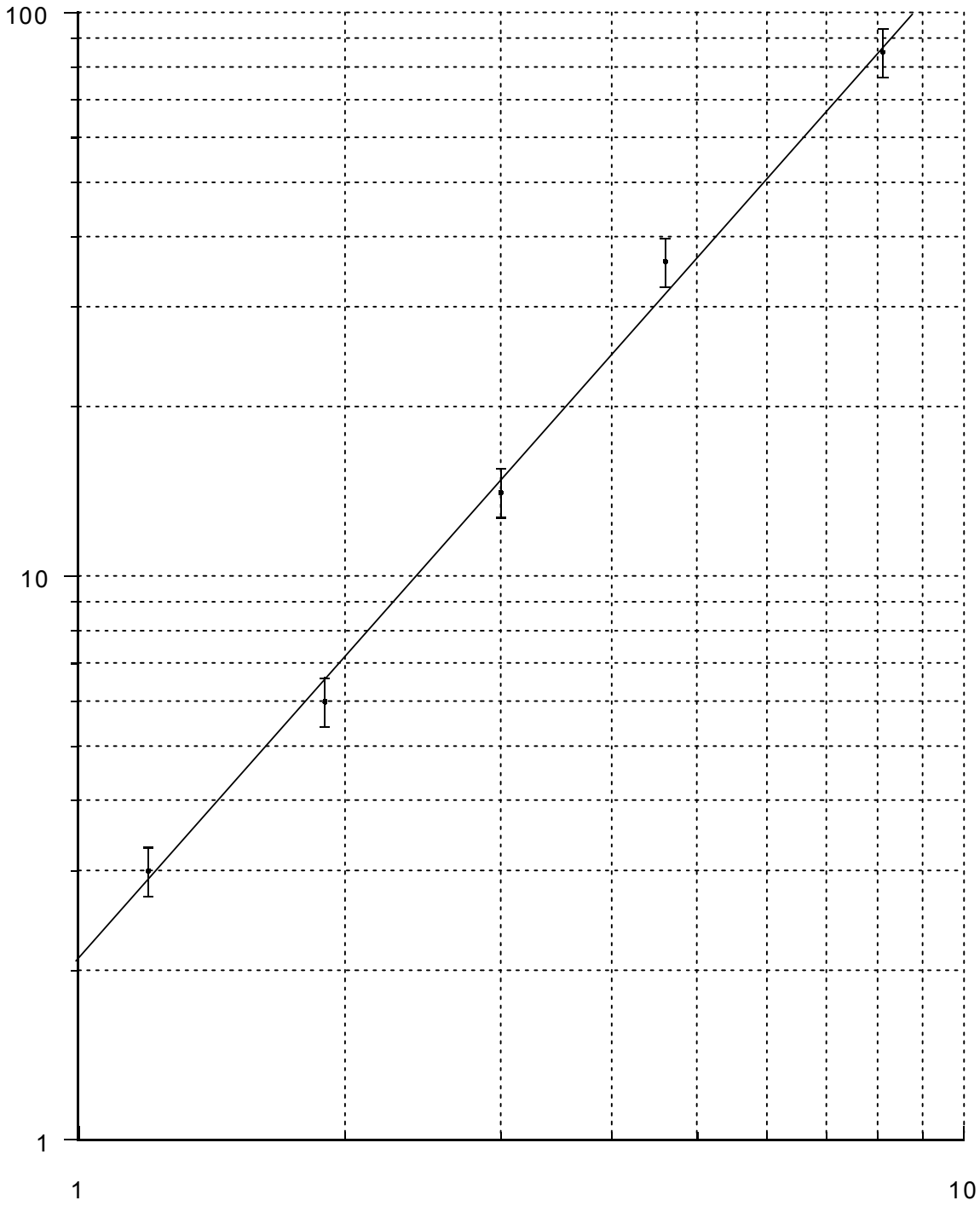
$$\ln(y) = \ln(b) + c\ln(x)$$

It remains to determine c and b from the graph. To do this, choose two points (x_1, y_1) and (x_2, y_2) on the best-fit line. Then $y_1 = b\,x_1^c$ and $y_2 = b\,x_2^c$

Taking the ratio of the two $\dfrac{y_1}{y_2} = \left(\dfrac{x_1}{x_2}\right)^c$

or $\ln\left(\dfrac{y_1}{y_2}\right) = c\ln\left(\dfrac{x_1}{x_2}\right)$ thus $\boxed{c = \dfrac{\ln(y_1) - \ln(y_2)}{\ln(x_1) - \ln(x_2)}}$

Graph 4 LogLog



From Graph 4 choosing data pairs

$$x_1 = 1.0,\ y_1 = 2.1$$
$$x_2 = 8,\quad y_2 = 85$$

gives

$$c = \frac{\ln(y_1) - \ln(y_2)}{\ln(x_1) - \ln(x_2)} = \frac{\ln(0.0247)}{\ln(0.125)} = 1.78$$

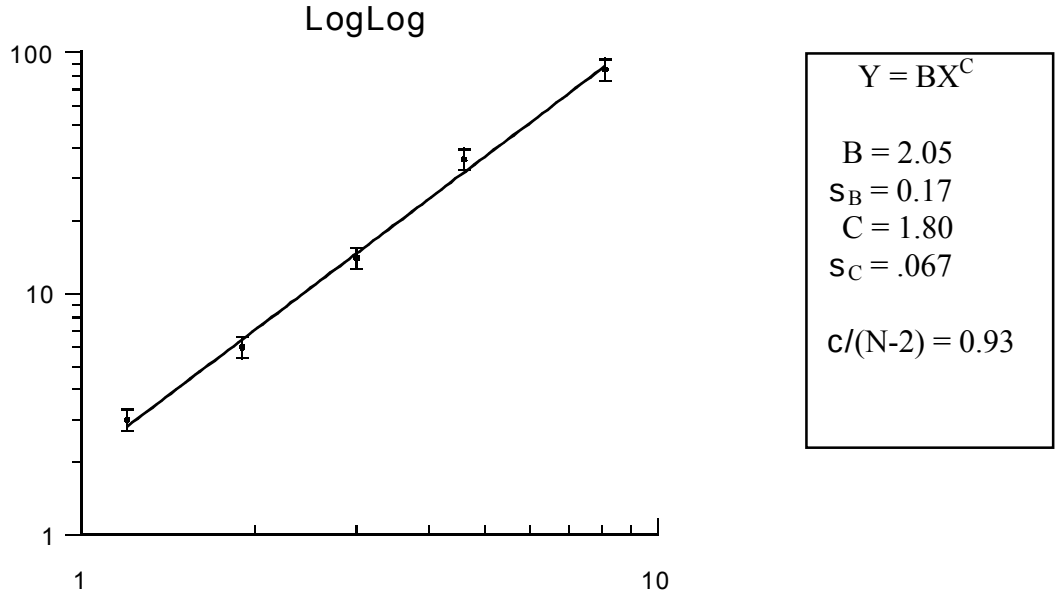
and $\quad b = \frac{y_1}{x_1^c} = \frac{2.1}{1.0^{1.78}} = 2.10 \qquad$ or $\qquad b = \frac{y_2}{x_2^{1.78}} = 2.098$

To check the results calculate y for x_2

$$y = (2.1)(8)^{1.78} = 85.0$$

which is in reasonable agreement with y_2

For comparison a Weighted LSF (See Chapter 4) gives



Note: Log plots compress the data diplay; for an example of the insight gained from difference plots for log-log plots see Example 2 page 5-17.

Semi-Logarithmic Graph Paper:

Another type of relationship which occurs quite frequently is of the form $w = me^{nx}$. Again, take the log of both sides of this expression;

$$\ln(w) = \ln(me^{nx}) = \ln(m) + nx$$

This expression is also of the form $y = a + nx$. In order to get a linear plot of data described by this relationship, we must have graph paper that has one axis ruled linearly and the other ruled logarithmically. Semi-log paper fits these requirements. As an example consider the following experimental data measuring the half life of Ag^{108}: where $C = C_0e^{-\lambda t}$, or $\ln(C) = \ln(C_0) - \lambda t$.

Time{sec)	Counts/8sec	s_C
236	1999	45
276	1567	40
316	1310	37
356	1130	34
396	930	31
436	751	28
476	557	25
516	488	23
556	412	21
596	310	19
636	283	18

Graph 5 is a plot of the data on semi-log paper. Since the range of C is two orders of magnitude ($10^2 - 10^3$, $10^3 - 10^4$), two-cycle paper was used. As in the previous case, the paper makes it possible to determine C_0 and λ. To do this choose two points (C_1, t_1) and (C_2, t_2) on the best-fit line. Then

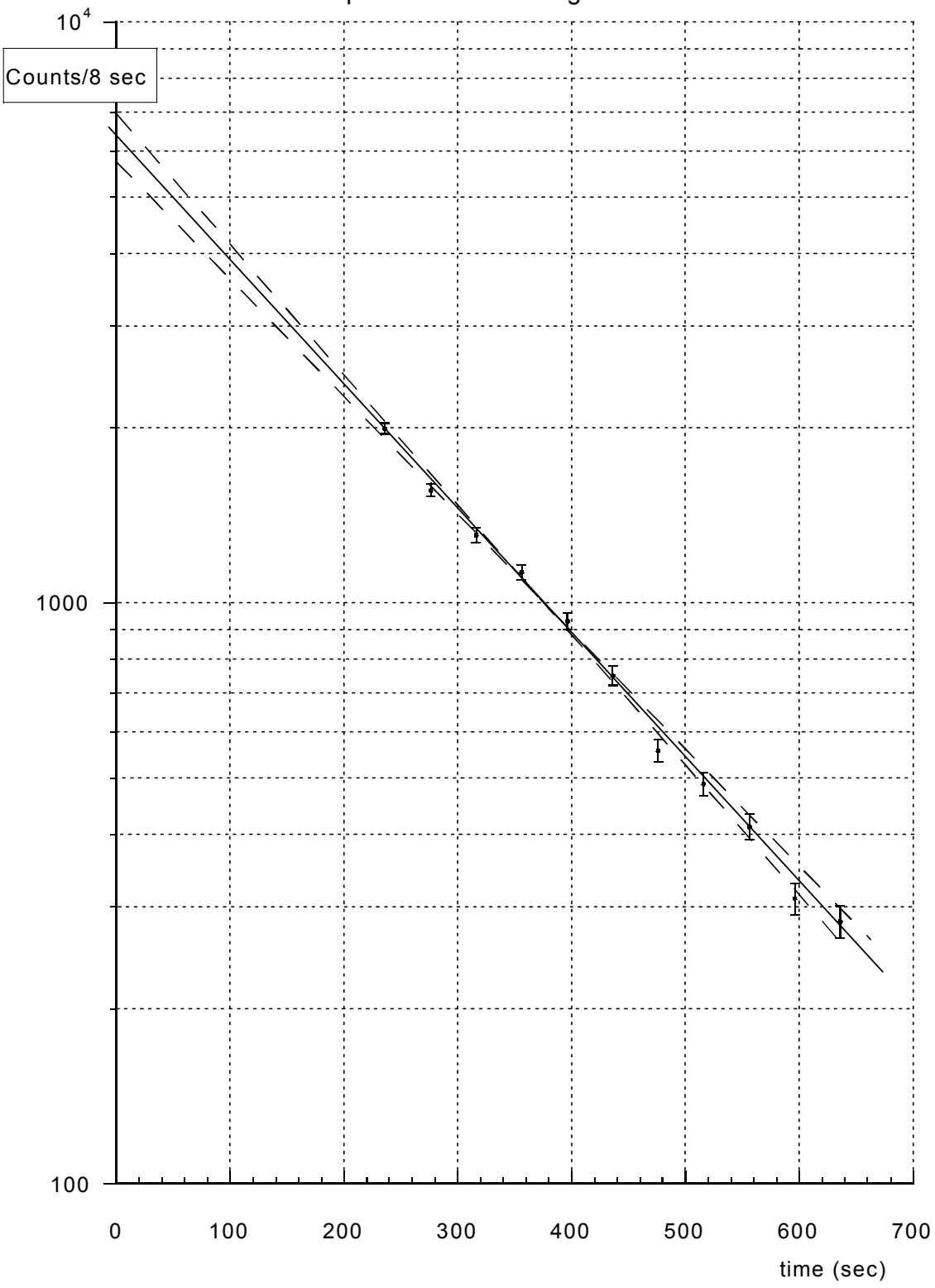
$$\ln(C_1) = \ln(C_0) - \lambda t_1$$

$$\ln(C_2) = \ln(C_0) - \lambda t_2$$

Subtracting the two equations gives

$$\boxed{\lambda = \frac{\ln(C_1) - \ln(C_2)}{t_2 - t_1}}$$

Graph 5 SemiLog



Take two points, widely separated, on the best-fit line

$C_1 = 2000$, $t_1 = 235$ sec; and $C_2 = 300$, $t_2 = 620$

$$\text{Slope} = -\lambda = -\left[\frac{\ln(2000) - \ln(300)}{620 - 235}\right] = -.00493 / \text{sec}$$

The intercept, C_0, can be estimated directly from the graph. From the slope and one point on the best-fit line (300 counts/8sec, 620 sec) C_0 may be determined.

$C_0 = Ce^{\lambda t} = (300)\exp[(0.00493)(620)]$

$C_0 = 6376$ counts/8 sec

From the dashed lines on Graph 5, an estimate of the uncertainty in λ and C_0 may be made.

$$\sigma_\lambda \cong \tfrac{1}{2}(\text{max. slope} - \text{min. slope}) \cong \tfrac{1}{2}(.00511 - .00464) = .0002$$

$$\sigma_{C_0} \cong \tfrac{1}{2}(\text{max.} - \text{min. intercept}) \cong \tfrac{1}{2}(7000 - 5700) = 650$$

Thus

$\lambda = -0.0049 \pm 0.0002$ /sec or better still $(-4.9 \pm 0.2) \times 10^{-3}$/sec

$C_0 = 6376 \pm 650$ counts/8 sec

NOTE: The log of any quantity is unitless. This follows from the fact that the exponent is always unitless. In $C = C_0 e^{-\lambda t}$, λt is unitless, dictating that $\lambda = 1/\text{sec}$ if $t =$ sec. Taking the log of $C = C_0 e^{-\lambda t}$ gives $\ln(C_1) = \ln(C_0) - \lambda t_1$. If λt is unitless then $\ln(C_0)$ must be unitless as well as $\ln(C_1)$. The equal sign applies to <u>units</u> as well as numbers.

Semi-log paper is also useful for plotting data taken over a large range of values. For instance, the response of an RC circuit may he measured over a range of frequencies from 1 kHz to 100 kHz. If the frequency is plotted on the log axis of say, 3 cycle semi-log paper, the full range of the data can be properly displayed on a single graph.

Log Difference Plots

The quantity

$$\frac{\ln(y_f) - \ln(y_i)}{\sigma_i / y_i} = \frac{\ln\left(y_f / y_i\right)}{\sigma_i / y_i}$$

Letting $y_f = y_i$ Ò Dy and using ln(1+ e) = e we get

$$\frac{\ln\left(1 \pm \Delta y / y_i\right)}{\sigma_i / y_i} = \frac{\pm \Delta y / y_i}{\sigma_i / y_i} = \frac{\pm \Delta y}{\sigma_i} = \frac{y_f - y_i}{\sigma_i}$$ as the quantity to be plotted.

Thus it is seen that for small differences, a log difference plot may be approximated by a linear difference plot.

Histograms

Graph 6 shows the frequency of occurrence f_i , of counts/l0 sec, n_i. A plot of frequencies, called a histogram, is often useful in establishing the range of variation of a quantity and the manner in which it varies (e.g. Gaussian, flat, etc.). The choice of bin width is critical to the appearance of a histogram and therefore critical to your ability to judge its shape. If the data is thought to be Gaussian or Poisson, a good initial choice of bin width is s/3.

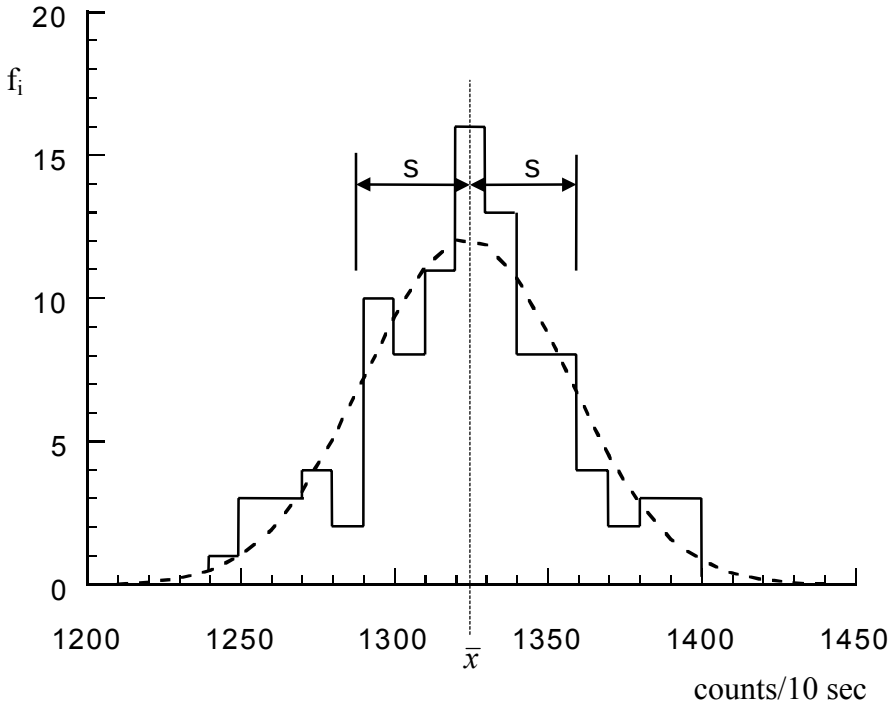
Initial data for Graph 6 was n_i = 1320, 1327, and 1293 counts/l0sec. Thus for Poisson statistics

$$\sigma = \sqrt{\bar{x}} \cong \sqrt{1300} \cong 36$$

Therefore choose a bin width of s/3 Ü10

GRAPH 6

Histogram of 100 ten second count intervals



NOTE: About 70 count intervals are within Ò 1s of $\bar{x}$

Analysis of Histogram Data

The statistical nature of the counting data should be reflected in the shape of the histogram, Graph 6. If the data is statistical, we can estimate a mean $\bar{x}$ and a standard deviation s, and fit a Gaussian curve to the histogram.

$$\bar{x} = \frac{\sum f_i n_i}{\sum f_i} = \frac{\sum f_i n_i}{N} \qquad (f_i = \text{number of count intervals in each bin})$$

Using the mid-bin value for n_i

$$\bar{x} = \frac{1(1245) + 3(1255) + 3(1265) + \ldots}{100} = 1323.6 \text{ counts} / 10 \text{ sec}$$

For Poisson statistics

$$\underline{\underline{\sigma = \sqrt{\bar{x}} \cong 36}}$$

$$\left\{ \begin{array}{l} \text{As a check} - \text{Gaussian statistics gives} \\ \sigma^2 = \dfrac{\sum f_i (n_i - \bar{x})^2}{N - 1} = 1101, \quad \text{thus } \sigma = 33 \end{array} \right\}$$

Fitting a Gaussian

$$y = \frac{\text{Area}}{\sigma\sqrt{2\pi}} \exp - (n - \bar{x})^2 / 2\sigma^2 \qquad \text{where Area} = N \text{ (bin width)} \qquad (1)$$

$$= 100(10) = 1000$$

See the curve, with $\bar{x}$ = 1323.6 and s = 36, superimposed on the histogram - Graph 6.

Estimating Sigma from the Histogram

From the histogram itself, one can estimate sigma from the full width at half maximum **(FWHM).**

From Equation (1) above, at $n = \bar{x}$ $\quad y = \dfrac{\text{Area}}{\sigma\sqrt{2\pi}} = P$ $\qquad$ (P = amplitude at the peak)

Thus at 1/2 maximum amplitude $y = P/2 = P \exp - \left[\left(x_{1/2} - \bar{x} \right)^2 / 2\sigma^2 \right]$

Giving $\left(x_{1/2} - \bar{x} \right)^2 = 2\sigma^2 \ln(2)$ $\quad$ now $\quad$ **FWHM** $= 2\left(x_{1/2} - \bar{x} \right)$

Thus $\qquad \boxed{\textbf{FWHM} = 2\sigma\sqrt{2\ln(2)} = 2.35\sigma}$

From the histogram - Graph 6, P is estimated to be 12; **FWHM** at y = 6 is 8(10);

Giving $\underline{s = \textbf{FWHM} / 2.35 = 80 / 2.35 = 34}$

A Comment on the Uneven Look of the Distribution in Graph 6

Are the variations in adjacent bins of the histogram to be expected? The sigma on the counts in an individual bin is the square root of the number of counts in that bin. For example, take two adjacent bins which should have approximately the same number of counts in them

bin(1310-1320) = 11 counts

bin(1330-1340) = 13 counts

The average number of counts is $12 \pm \sqrt{12}$.

Thus the bin in the middle is within statistics as it is less than $2s$ from an expected 12 or 13 counts.

Some Final Comments on Uncertainties and Error Bars

Rarely in actual research data will the functional relationship between measured quantities be a foregone conclusion, nor will the complete range of physical quantities impacting the result be fully anticipated. In these cases a data plot without error bars is totally inadequate for assessing a best-fit form to the data. In spite of the fact that the above comments do not pertain to most of what you will do in the lab, the emphasis placed on the seemingly superfluous and sometimes tedious aspects of data analysis is very useful. You must use these more obvious examples generated in the lab experiments to develop data analysis techniques and to sharpen your insight into the range of conclusions possible, against the day when very little will be obvious; and that data point several sigma from its expected position may be the first measurement of something new.

The history of science contains many examples of the measurement of new unexpected effects that were overlooked in the data. The Josephson Junction Effect is a more recent example that comes to mind.

Rules of Graphing

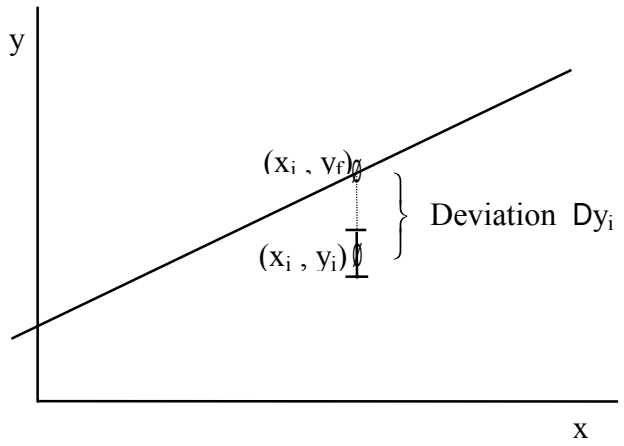
1. Whenever possible either plot as you are taking data, or immediately after, as gaps in the data can easily be seen and filled in.

2. Use appropriately labeled scales that will display the data to best advantage; namely, scale the data plot so the entire graphing surface is used.

3. The criteria of a fit line - one that when drawn through the data has approximately as many points on one side of the line as the other.

4. Never stop graphing until you see the error bars.

Least Squares Fits (LSF)



Weighted LSF

For a set of data bearing a linear relationship there exists a description of that relationship in the form $y_f = a + b\,x_i$

If data exists in the form x_i , y_i Ò s_i (s_i being derived from the assumed Gaussian distribution of the values of y_i at a given value of x_i) the probability of making the observed measurement y_i is

$$P_i = \frac{1}{\sigma_i\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{y_i - y_f}{\sigma_i}\right)^2$$

The probability of making the observed set of measurements of the N values of y_i is the product of the individual probabilities

$$P_{a,b} = \prod_{i=1}^{N} P_i = \left(\prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}}\right) \exp\left\{\sum -\frac{1}{2}\left(\frac{\Delta y_i}{\sigma_i}\right)^2\right\} \qquad (\Delta y_i = y_i - y_f)$$

Therefore the best estimates for a and b are those that maximize $P_{a,b}$ or equivalently minimize the sum in the exponential.

Weighted LSF cont.

Define

$$\chi^2 \equiv \sum\left(\frac{y_i - y_f}{\sigma_i}\right)^2 = \sum\left(\frac{1}{\sigma_i^2}(y_i - a - b\,x_i)^2\right) \tag{1}$$

Now minimize c^2 by taking

$$\frac{\partial\chi^2}{\partial a} = -2\sum\left(\frac{1}{\sigma_i^2}(y_i - a - b\,x_i)\right) = 0$$

and

$$\frac{\partial\chi^2}{\partial b} = -2\sum\left(\frac{x_i}{\sigma_i^2}(y_i - a - b\,x_i)\right) = 0 \tag{2}$$

yielding the LSF equations

$$a = \frac{1}{\Delta}\left(\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2}\right) \tag{3}$$

and

$$b = \frac{1}{\Delta}\left(\sum\frac{1}{\sigma_i^2}\sum\frac{x_i\,y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2}\right) \tag{4}$$

where

$$\Delta = \sum\frac{1}{\sigma_i^2}\sum\frac{x_i^2}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2 \tag{5}$$

NOTE: From Equation (2)

$$b = \frac{\sum x_i y_i/\sigma_i^2 - a\sum x_i/\sigma_i^2}{\sum x_i^2/\sigma_i^2}$$

Due to round off errors avoid the product, $a\Sigma 1/s_i^2$, particularly near $a = 0$. This product results from solving for b in the equation prior to Equation (2).

Caution: Before using these LSF equations on non-linear forms see page 4-11.

Uncertainty of Fit Parameters (σ_a and σ_b)

The calculation of the uncertainties in the parameters *a* and *b* follows from the general equation, Equation (4) page 2 - 2.

$$\sigma_n^2 = \sum \left(\sigma_i^2 \left(\frac{\partial n}{\partial y_i} \right)^2 \right) \tag{6}$$

From the most general equations for *a* and *b* (Equations (3) and (4)) we get

$$\frac{\partial a}{\partial y_i} = \frac{1}{\Delta} \left(\frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \frac{x_i}{\sigma_i^2} \sum \frac{x_i}{\sigma_i^2} \right) \tag{7}$$

$$\frac{\partial b}{\partial y_i} = \frac{1}{\Delta} \left(\frac{x_i}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^2} \sum \frac{x_i}{\sigma_i^2} \right) \tag{8}$$

giving

$$\boxed{\sigma_a^2 \cong \frac{1}{\Delta} \sum \frac{x_i^2}{\sigma_i^2} \qquad \text{and} \qquad \sigma_b^2 \cong \frac{1}{\Delta} \sum \frac{1}{\sigma_i^2}} \tag{9}$$

where D is given by Equation (5).

NOTE: The above sigmas depend only on the uncertainties of the individual data points and are not affected by the scatter in the data. This could cause misleading results if $c^2/(N\text{-}2)$ is not calculated. See page 4-5 and the Example page 4-8.

Unweighted Fit (or Equal Weights)

In the case where $s_i = s$ (i.e. all the s_i's are equal) some simplification results,

$$a = \frac{1}{\Delta'}\left(\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i\right) \tag{10}$$

$$b = \frac{1}{\Delta'}\left(N \sum x_i y_i - \sum x_i \sum y_i\right) \tag{11}$$

where

$$\Delta' = N \sum x_i^2 - \left(\sum x_i\right)^2 \tag{12}$$

Note that the equations for a and b are now independent of s and may be derived from Equations (3) and (4) by the substitution of $s_i = s = 1$, where $S1/s_i^2 = N$ with this substitution.

Uncertainties in the Unweighted Case

In the case where $s_i = s$ the quantities of Equation (9) simplify to

$$\sigma_a^2 \cong \frac{\sigma^2}{\Delta'} \sum x_i^2 \qquad \text{and} \qquad \sigma_b^2 \cong N \frac{\sigma^2}{\Delta'} \tag{13}$$

where D' is given by Equation (12).

If s is known, the uncertainties may be calculated from these equations. If s is not known, an estimate of it may be made from the scatter of the data by calculating a quantity s^2, known as the variance.

$$\sigma^2 \cong s^2 = \frac{1}{N-2} \sum (y_i - y_f)^2 \tag{14}$$

$$= \frac{1}{N-2} \sum \left(y_i - a - bx_i\right)^2$$

$$s^2 \cong \frac{1}{N-2}\left(\sum y_i^2 - a \sum y_i - b \sum x_i y_i\right)$$

Even if s is known, it is a good idea to calculate the variance, as a large difference between the two may suggest that the data requires further scrutiny.

χ^2 Test for a Weighted L.S.F.

The note under Equation (9) implies that a data point several s off the best-fit line could go undetected if only the parameters of Equation (9) were calculated. To prevent this we need an additional quantity which is sensitive to the scatter in the data, similar to the variance s^2 in the unweighted fit,

Define the Reduced Chi^2 as

$$\boxed{\frac{\chi^2}{N-2} = \frac{1}{N-2}\sum \frac{1}{\sigma_i^2}\left(y_i - y_f\right)^2} \tag{15}$$

What then is a reasonable value of this quantity for a good fit? Define a good fit to say, 10 points, as 7 points within 1s of the fit line and 3 points within 2s (i.e. y_i - $y_f = 2s_i$)

$$\chi^2/N-2 = \frac{1}{8}\left[7(1)^2 + 3(2)^2\right] \cong 2 \qquad \text{(Ideal fit: } \chi^2/N\text{-}2 \cong 1\text{)}$$

A more useful form of Equation (15) for programming on a pocket calculator* is

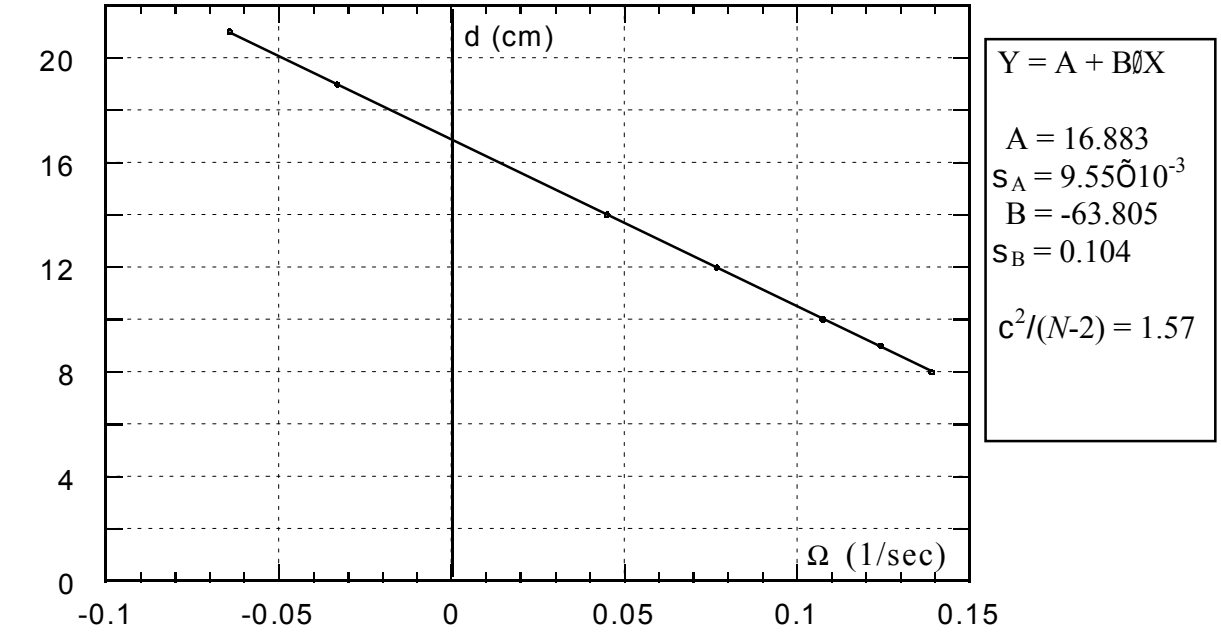
$$\frac{\chi^2}{N-2} = \frac{1}{N-2}\left\{\sum \frac{y_i^2}{\sigma_i^2} - a\sum \frac{y_i}{\sigma_i^2} - b\sum \frac{x_i y_i}{\sigma_i^2}\right\} \tag{16}$$

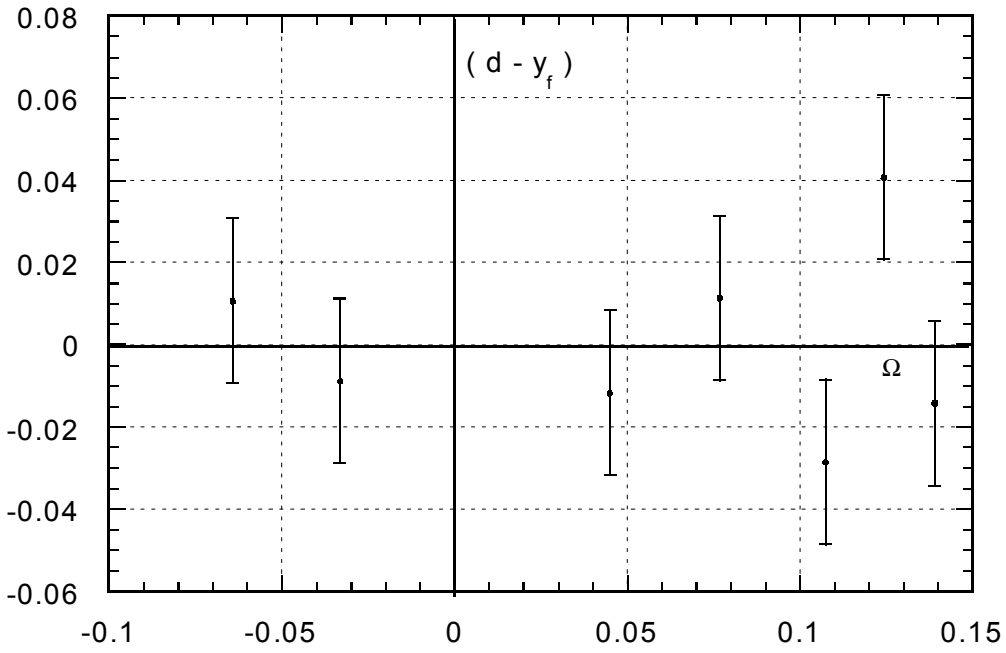
If the reduced c^2 is not around one, this indicates, a lack of consistency between the scatter in the data and the size of the error bars, or a point several sigma from the fit line, or the presence of some additional information in the data. In any case, a reduced c^2 not around one, suggests that the data needs a closer look. See the example on page 3-10 and the quadratic LSF to the same data on page 4-7.

---

* Caution: Equation (16) often leads to negative results when used on a computer with single precision (7significant figures) due to round off errors. Rather than going to double precision (14 significant figures) it is more efficient to use Equation (15) and a data array.

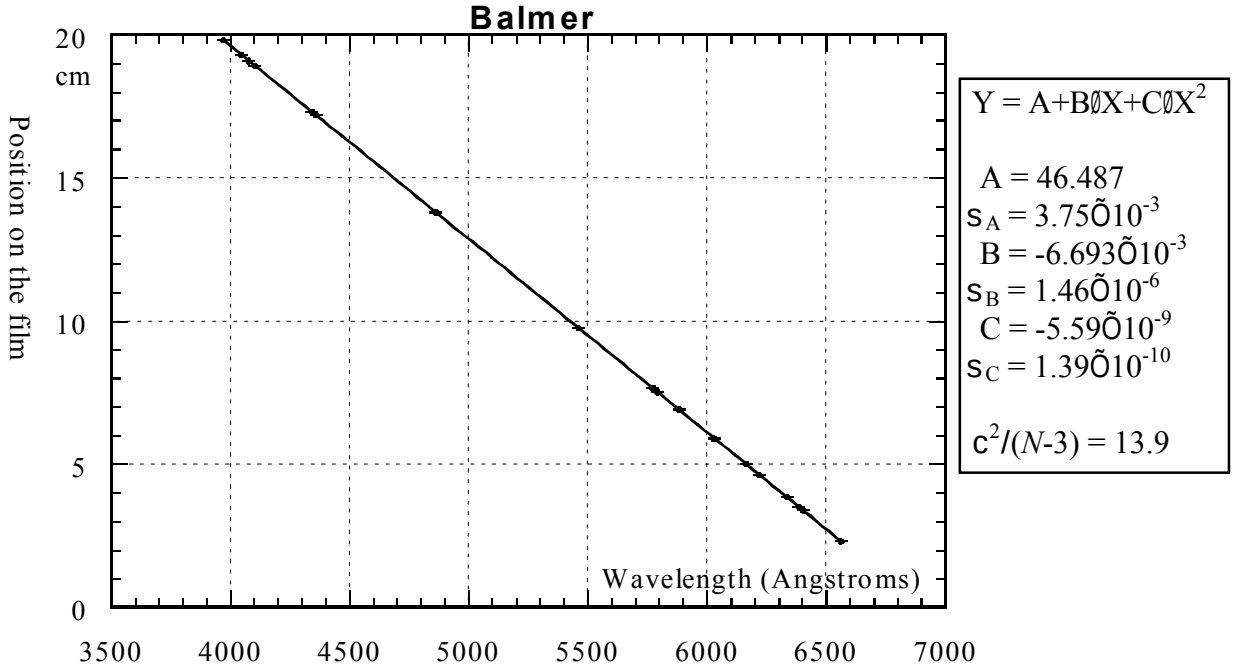
Weighted LSF Example – Maxwell Top Data (See page 3-5 for a listing of the data.)

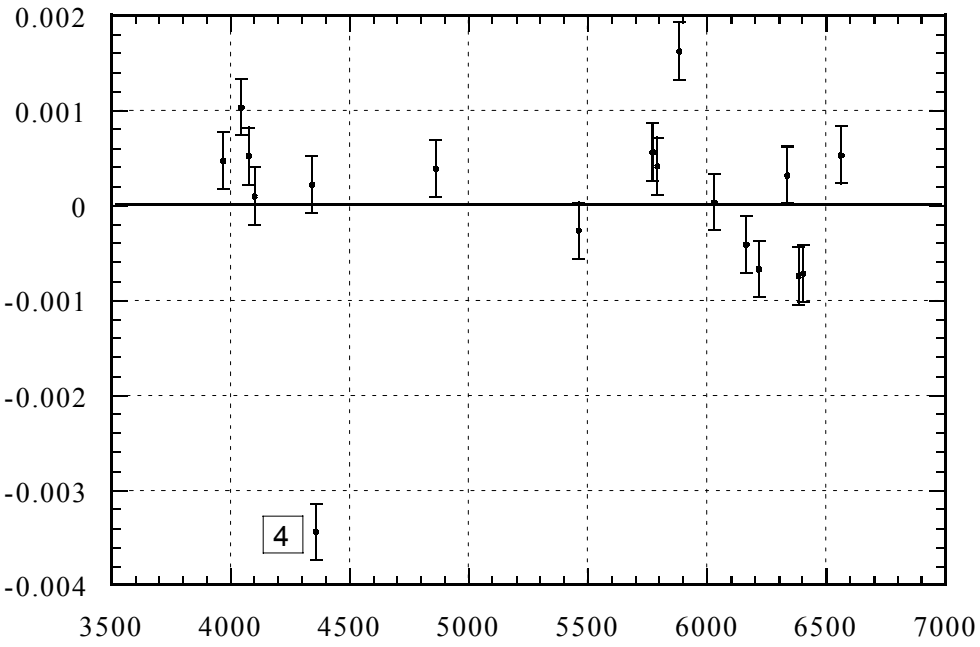




Compare these results with those obtained on pages 3-6 to 3-8.

Quadratic Fit to the Data Plotted on page 3-10





If item #4 is removed from the data set $c^2/(N-3) = 4.2$

Example

A linear fit to the data listed below is a dramatic example of the necessity of plotting data, and calculating the Reduced χ^2 in a weighted fit.

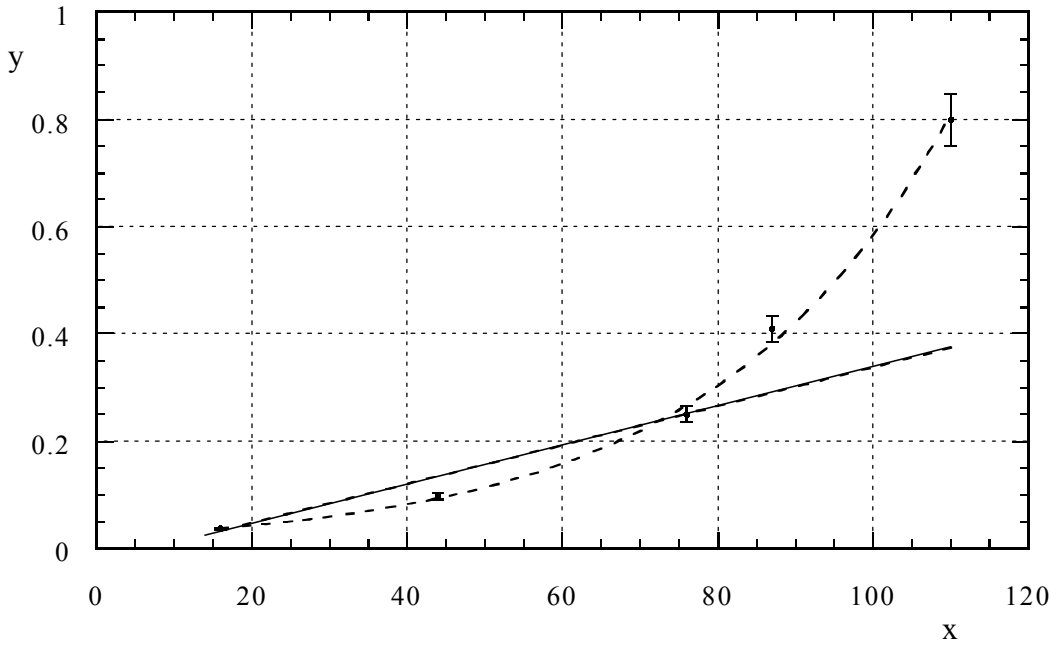
x	16	44	76	87	110	
y	0.037	0.09	0.25	0.4	0.80	s_y = Ò 6%

Without a plot of the data, an erroneous linear fit (y = a + b x) to this exponential data might be attempted and yield:

$a = (-2.45 \pm 0.37) \times 10^{-2}$, $b = (3.62 \pm 0.15) \times 10^{-3}$, $\chi^2/N\text{-}2 = 49$

Note that the uncertainty on the slope, *b*, is only 4% and would not necessarily alert you to the problem with the fit. On the other hand, $\chi^2/N\text{-}2 = 49$.

The Reduced χ^2 is definitely alerting you to a problem in the fit. The data and the linear fit line are plotted on the graph below. It is clear from this example that the uncertainties on the slope and intercept are meaningful only if the Reduced $\chi^2 \approx 1$.



Comment on the Variance in the Unweighted Case

Note the similarity of Equations (14) and (15). The quantity calculated in Equation (14) is the sigma necessary on each data point for an ideal fit.

To generate Eq. (14), set Eq. (15) equal to one, and all σ_i's = σ. This gives

$$1 = \frac{1}{N-2}\left(\frac{1}{\sigma^2}\right)\Sigma(y_i - y_f)^2$$

which leads to Eq. (14) when solved for σ^2.

To illustrate the point do an unweighted fit to the Maxwell Top data of Example 1. This fit gives σ = 0.025. Now do a weighted fit to this data setting all σ_d = 0.025 and note that c^2/N-2 = 1.

Thus Unweighted fits do not provide a test of the data, as they assume a perfect fit and produce results based on that assumption, with no check.

Computer Round off Errors

To avoid round off errors in a Least Squares computer program, resulting from small differences generated by subtracting large sums, use double and triple sums. These have the advantage of taking a difference at each term in the sum thus avoiding this problem associated with small differences of large numbers.

Take for example equation (3) page 4-2

$$a\Delta = \sum \frac{x_i^2}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2}$$

can be written as

$$a\Delta = \sum_k \sum_i \frac{1}{\sigma_k^2 \sigma_i^2}\left(x_k^2 y_i\right) - \sum_k \sum_i \frac{1}{\sigma_k^2 \sigma_i^2}(x_k x_i y_i)$$

$$a\Delta = \sum_k \sum_i \frac{1}{\sigma_k^2 \sigma_i^2}\left(x_k^2 y_i - x_k x_i y_i\right) \quad = \quad \boxed{\sum_k \sum_i \frac{1}{\sigma_k^2 \sigma_i^2} x_k y_i (x_k - x_i)}$$

This method is most useful in quadratic fits where sums up to x^4 must be formed. The generalization to the triple sums needed for a quadratic fit is obvious.

Weighted Fits to Other Functional Forms

Several other functional forms can be linearized. Take the equation

$z = me^{bx}$ (data taken in the form x_j, z_j Ò s_j)

Taking the log of both sides gives: $\ln z = \ln m + bx$; which is in the form: $y = a + bx$,

where $y = \ln z$, $a = \ln m$, and $b = b$.

Transforming the weighting factor s_j to s_i through Equation (4) page 2-2 gives

$$\sigma_i = \frac{dy}{dz_j}\sigma_j = \frac{d(\ln z_j)}{dz_j}\sigma_j = \frac{\sigma_j}{z_j}$$

Thus the data is entered in Equations (3), (4), (9), and (15) with

$$x_i = x_j, \quad y_i = \ln(z_j), \text{ and } s_i = s_j / z_j$$

The uncertainties s_a and s_b generated by Equation (9) are related to the uncertainties in m and b through Equation (6). In this case

$$\sigma_a = \frac{\partial(\ln m)}{\partial m}\sigma_m = \frac{\sigma_m}{m}, \quad \text{and} \quad \sigma_b = \sigma_b$$

Thus the results of Equations (3), (4), (9), and (15) are

$$\ln(m),\ b,\ s_m / m,\ s_b,\ \text{and}\ c^2/(N\text{-}2)$$

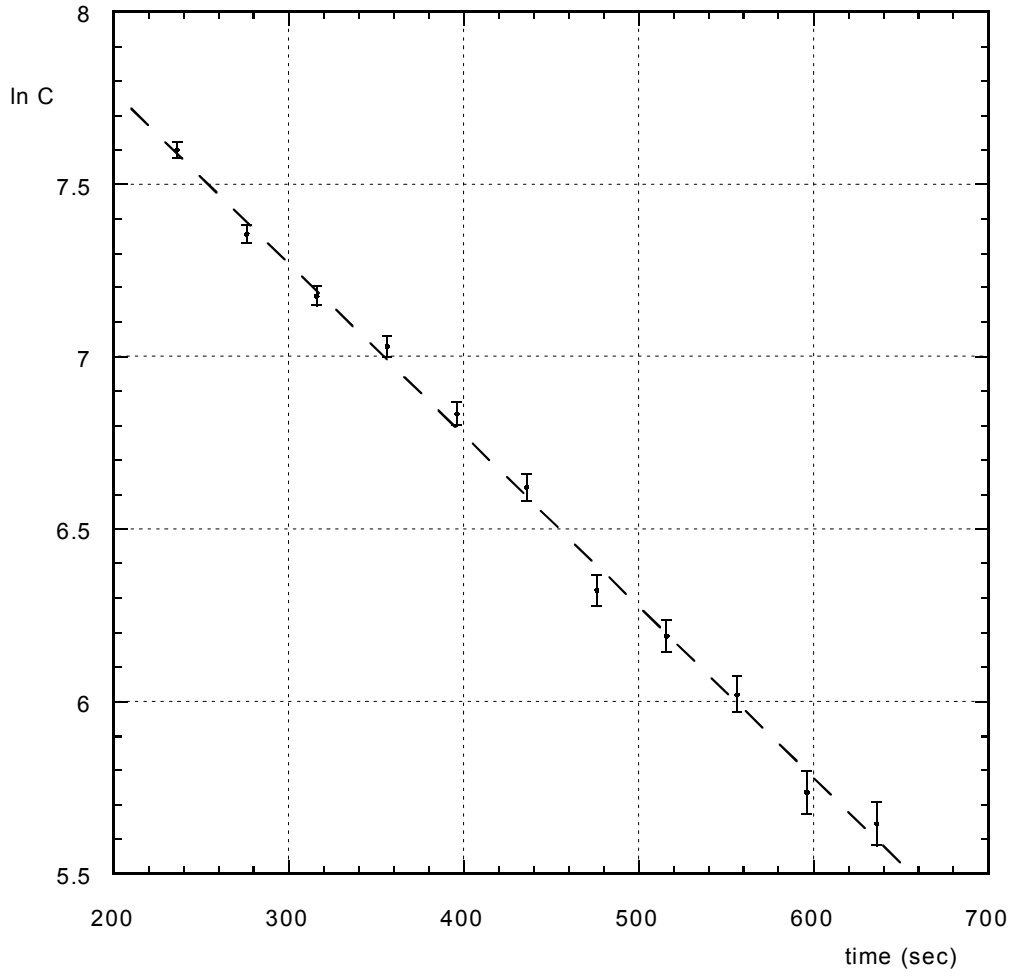
The necessity for tranforming the weighting factor s_j can be understood by plotting the data in the linear form in which it is presented to the LSF equations (i.e., plot linearly, the data on page 3-14, in the form ln C_i vs. t_i).

Data from Example 2 page 4-13

t_i	C_i	s_C	ln C_i	$s_i = s_C / C_i$
236	1999	45	7.60	0.022
276	1567	40	7.36	0.025
etc.				

The full data set is plotted on the graph on the next page.

A linear plot of the data from Example 2 page 4-12. Plotted points t_i , $\ln C_i \pm s_i$



Note the similar appearance of this plot and Graph 5, page 3-15, which is a semi-log plot of the same data.

It is obvious that the semi-log paper automatically takes the log of the dependent variable when you plot the data; but it is less obvious that it also transforms the uncertainty as well. The information lost in this transformed linearized fit is the asymmetry in the error bars visible at small values of y on Graph 5 page 3-15.

Example 2 - Silver Activation Experiment

The exponential dependence of the counts with time, in the activation of silver experiments, is related by the equation

$$C = C_0 e^{-\lambda t}$$

Data

Raw counts, *R*, accumulated in 8 sec. per channel.

Background, *B*, was 54Ò1 counts per 8 sec. (See Example page 1 - 14)

Data is statistical, i.e. $\sigma_R = \sqrt{R}$

Channel No.	Time (mid-bin)	*R*	$C = R-B$	$s^2_C = s^2_R + s^2_B$
30	236 sec.	2053	1999	45
35	276	1621	1567	40
40	316	1364	1310	37
45	356	1184	1130	34
50	396	984	930	31
55	436	805	751	28
60	476	611	557	25
65	516	542	488	23
70	556	466	412	21
75	596	364	310	19
80	636	337	283	18

Quantities entered in Equations (3), (4), (9), and (15) are

$$x_i = t_i\ ,\ y_i = \ln C_i\ ,\ s_i = s_C / C_i$$

Results are: $c^2/(N\text{-}2) = 1.12$

$\ln C_0 = 8.76$, $l = 0.00496/\text{sec}$, $s_{C_0}/C_0 = 0.035$, and $s_l = 9 \times 10^{-5}$

Thus $\underline{C_0 = 6374}$ counts and $\underline{\sigma_{C_o} = \pm 223}$

The half life of the Ag^{108} isotope is

$$\underline{t_{1/2} = \frac{0.693}{\lambda} = 139.7\,\text{sec}} \quad \text{and} \quad \sigma_{t_{1/2}} = \frac{0.693}{\lambda^2}\sigma_\lambda \quad \text{giving} \quad \underline{\sigma_{t_{1/2}} = \pm 2.5\,\text{sec}}$$

Compare these results with those obtained from Graph 5 page 3 - 16.

Extrapolated or Interpolated Value $\hat{y}$ and its Uncertainty $\sigma_{\hat{y}}$

$\hat{y} = a + bx$ (a and b previously determined from a LSF to a data set)

From Equation (2), page 2-2

$$\sigma_{\hat{y}}^2 = \left(\frac{\partial \hat{y}}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial \hat{y}}{\partial b}\right)^2 \sigma_b^2 + 2\sigma_{ab}\left|\frac{\partial \hat{y}}{\partial a}\right|\left|\frac{\partial \hat{y}}{\partial b}\right|$$

The covariant term must be included because a and b are not independent of each other.

Thus $$\boxed{\sigma_{\hat{y}}^2 = \sigma_a^2 + x^2\sigma_b^2 + 2x\sigma_{ab}}$$

For simplicity, take the unweighted case where s_a and s_b are given by Equation (13). Generalizing from Equation (6) define s_{ab} as

$$\sigma_{ab} = \sigma^2 \sum \left(\frac{\partial a}{\partial y_i}\right)\left(\frac{\partial b}{\partial y_i}\right) \qquad \text{[see Equations (7) and (8)]} \qquad (17)$$

$$= \frac{\sigma^2}{(\Delta')^2} \sum \left\{(\Sigma x_i^2 - x_i \Sigma x_i)(N x_i - \Sigma x_i)\right\}$$

$$= \frac{\sigma^2}{(\Delta')^2}\left((\Sigma x_i)^3 - N\,\Sigma x_i\,\Sigma x_i^2\right) = \frac{\sigma^2}{(\Delta')^2}(\Delta' \Sigma x_i)$$

$$\sigma_{ab} = \frac{-\sigma^2}{\Delta'}\Sigma x_i$$

Thus $$\boxed{\sigma_{\hat{y}}^2 = \frac{\sigma^2}{\Delta'}\left(\Sigma x_i^2 + N x^2 - 2x\Sigma x_i\right)} \qquad (18)$$

or $$\sigma_{\hat{y}}^2 = \sigma_a^2 + x^2\sigma_b^2 - (2x\frac{\sigma^2 \Sigma x_i}{\Delta'})$$

In the Weighted Case $$\boxed{\sigma_{\hat{y}}^2 = \frac{1}{\Delta}\left[\Sigma(x_i^2/\sigma_i^2) + x^2\Sigma(1/\sigma_i^2) - 2x\Sigma(x_i/\sigma_i^2)\right]} \qquad (19)$$

Note that at x = 0, Equation (18) becomes $\sigma_{\hat{y}}^2 = \frac{\sigma^2}{\Delta'}\Sigma x_i^2$ which is identical to the sigma of the intercept σ_a^2 in Equation (13).

Note also: at $x = \bar{x} = (1/N)\Sigma x_i$, Equation (18) gives $\sigma_{\hat{y}} = \sigma/\sqrt{N}$

Example 3 - Silver Activation Experiment

The neutron activation of silver experiment provides an application of the previous section. The counts from the short lived Ag^{110} isotope are mixed with the longer lived Ag^{108} counts in the early channels, and it is therefore necessary to extrapolate the Ag^{108} fit of Example 2 to estimate an Ag^{108} contribution to the early channels. Take for instance the data of channel 4: mid-bin time 28 sec., R = 15860 counts/8sec. How much of R is due to the Ag^{108} isotope?

Using the results of Example 2

$$\ln C = \ln C_0 - \lambda t = \ln(6374) - 0.00496(28) = 8.621$$

$$C = 5547 \text{ counts/8 sec}$$

The uncertainty in C is related to $\sigma_{\hat{y}}$ of Equation (19) through Equation (6)

$$\sigma_{\hat{y}} = \frac{\partial(\ln C)}{\partial C}\sigma_C = \frac{\sigma_C}{C}$$

Using the sums accumulated in Example 2, in Equation (19) gives: $\sigma_{\hat{y}} = 0.03284$

Thus $s_C = (5547)\,(0.03284) = \underline{182}$

The Ag^{108} contribution to R at 28 sec. is thus estimated as

(5547 Ò182)/8sec

To complete the estimate of the Ag^{110} data point at 28 sec, several more adjustments of R must be made.

Example 3 (cont.)

Dead Time Correction:

If the count rate is high enough the counting apparatus may miss a legitimate count while processing a previous count.

The actual rate R' in terms of the measured rate, R = 15,860/8 sec, and the dead time, t_d , is

$$R' = \frac{R}{1 - R\,t_d} \qquad t_d = (0.9 \pm 0.1)\text{x}10^{-6} \text{ (8sec)}$$

$$R' = 16089 \text{ counts/8sec}$$

By applying Equation (6) to the equation for R'

$$\sigma_{R'}^2 = \frac{\sigma_R^2 + R^4\sigma_{t_d}^2}{(1 - R\,t_d)^4} = 17469$$

Thus

$$R' = (16089 \pm 132) \text{ counts/8 sec}$$

Subtracting the background rate, B = (54 ± 1) counts/8 sec gives

$$R'' = R' - B = 16035 \pm 132 \qquad (\sigma_{R''}^2 = \sigma_{R'}^2 + 1)$$

Finally $\text{Ag}^{110} = R'' - \text{Ag}^{108} = 10488$ counts/8sec.

and $$\sigma_{110}^2 = \sigma_{R''}^2 + \sigma_{108}^2 = (132)^2 + (182)^2$$

Giving

$$\boxed{\text{Ag}^{110} = (10488 \pm 225) \text{ counts/8 sec}}$$

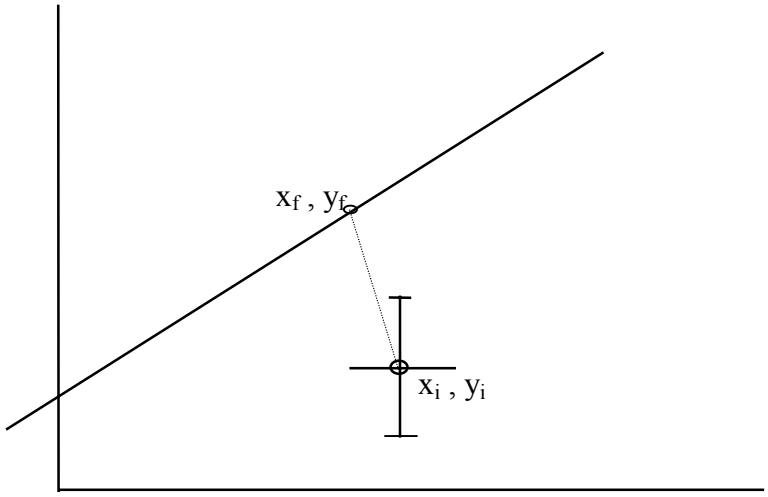
Dealing with Uncertainties in Both x And y

If the data is in the form (x Ò s_x , y Ò s_y) Demming[1] has shown that the proper form for c^2 is

$$\chi^2 = \sum \left(\frac{(x_i - x_f)^2}{\sigma_x^2} + \frac{(y_i - y_f)^2}{\sigma_y^2} \right) \tag{20}$$

Setting the appropriate partial derivatives to zero results in a set of equations that cannot be solved in closed form and must be applied iteratively. A useful form of these equations is given by Williamson in the Canadian Journal of Physics, 46,1845 (1968).

The form of c^2 given in Equation (20) minimizes the weighted distance from data point (x_i, y_i) to the fit line at point (x_f , y_f).



This form of c^2 goes to the form of Equation (1) as s_x and also (x_i - x_f) goes to zero; and again the distance minimized is the vertical distance.

[1] Demming- Statistical Adjustment of Data; Wiley (1943) (also available from Dover Press) (See also - A Better Least-Squares Method when Both Variables have Uncertainties - Matthew Lybanon - Am. J. Physics 52(1), Jan.1984)

Effective Variance

A quantity called the effective variance and defined for linear fits as

$$\sigma_i^2 = \sigma y_i^2 + b_n^2 \, \sigma x_i^2$$

is a way of including the effect of the uncertainty in x (σx_i) in the standard least squares Equations (3), (4), and (5). J. Orear has shown that the effective variance is more generally stated as

$$\boxed{\sigma_i^2 = \sigma y_i^2 + \left(\frac{\partial f}{\partial x}\right)_n^2 \sigma x_i^2} \qquad (21)$$

and is exact for linear fits. (See Appendix)

An iterative algorithm that uses Equations (3), (4), and (5) and the effective variance is as follows:

(1) Do a weighted LSF (data in the form x_i , y_i Ò $s y_i$) to get an estimate of the slope, b_1.

(2) Calculate the effective variance for each data point. $\sigma_i = \sqrt{\sigma y_i^2 + b_1^2 \sigma x_i^2}$

(3) Do a weighted LSF using s_i (data in the form x_i , y_i Ò s_i) and obtain b_2.

(4) If b_2 differs from b_1 , redo steps (2) and (3) and compare b_3 and b_2.

The algorithm will generally converge in 3 iterations.

A good test of whichever fitting routine is used is to swap the variables and check the invariance of the results. Thus one should check to see if

$$b_{xy} = \frac{1}{b_{yx}} \qquad \text{and} \qquad a_{xy} = -\frac{a_{yx}}{b_{yx}}$$

Finally:

All the ideas in these first four chapters are implemented in the program **PFIT45.EXE** available on the WEB at - www.pma.caltech.edu/~dons/ph3-7/

[2] J. Orear - CLNS-81/496 June, Cornell University, Ithaca, N.Y.

Pearson's data listed below is a good test of any fitting algorithm.

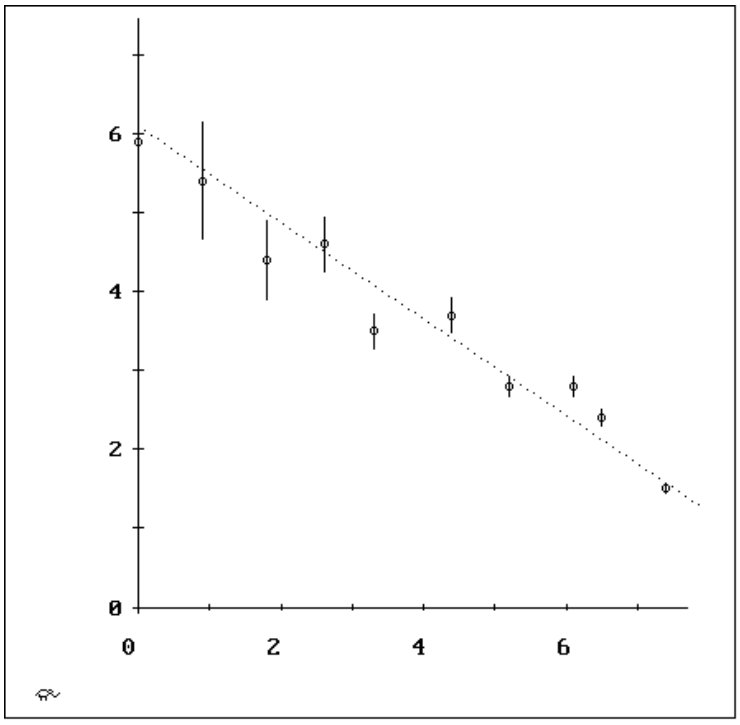
x	s_x	y	s_y
0.0	0.03	5.9	1.0
0.9	.03	5.4	.74
1.8	.04	4.4	.5
2.6	.035	4.6	.35
3.3	.07	3.5	.22
4.4	.11	3.7	.22
5.2	.13	2.8	.12
6.1	.22	2.8	.12
6.5	.74	2.4	.10
7.4	1.0	1.5	.04

Williamson's results for this data are invariant and equal to:

$$a = 5.45, \quad b = -0.480 \pm 0.057$$

The results obtained using the effective variance and the standard least squares equations are listed with the graphs that follow.

Initial fit to x, y ± s_y

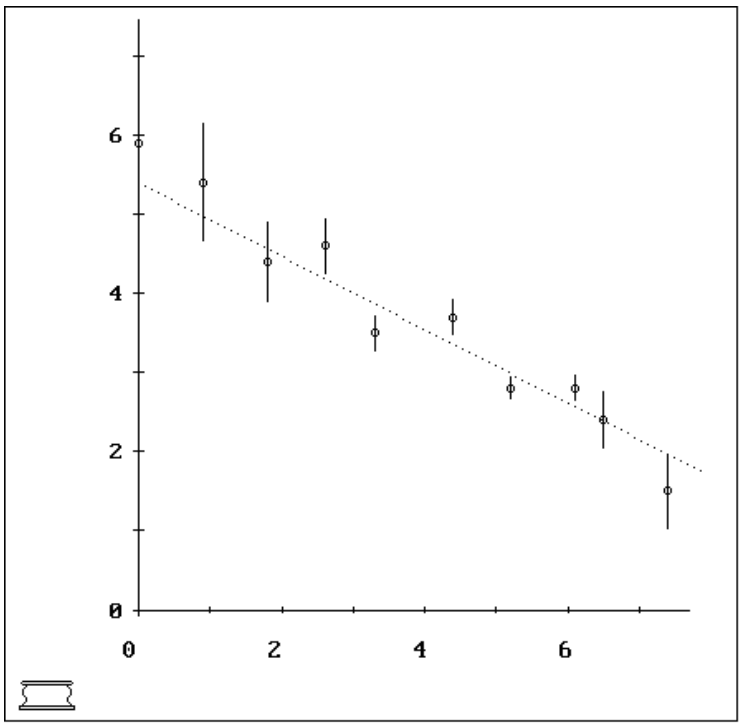



```
sigxy

 XOFF =  0
   DX =  1
 YOFF =  0
   DY =  1

  Y = A+B*X
 (N = 10 )
     A =  6.098231
SIG(A) =  .2030029
     B = -.6108411
SIG(B) =  2.980093E-02
CHI^2/(N-2) =
          4.327601
```

The 3rd iteration plotted with the effective variance s_i .

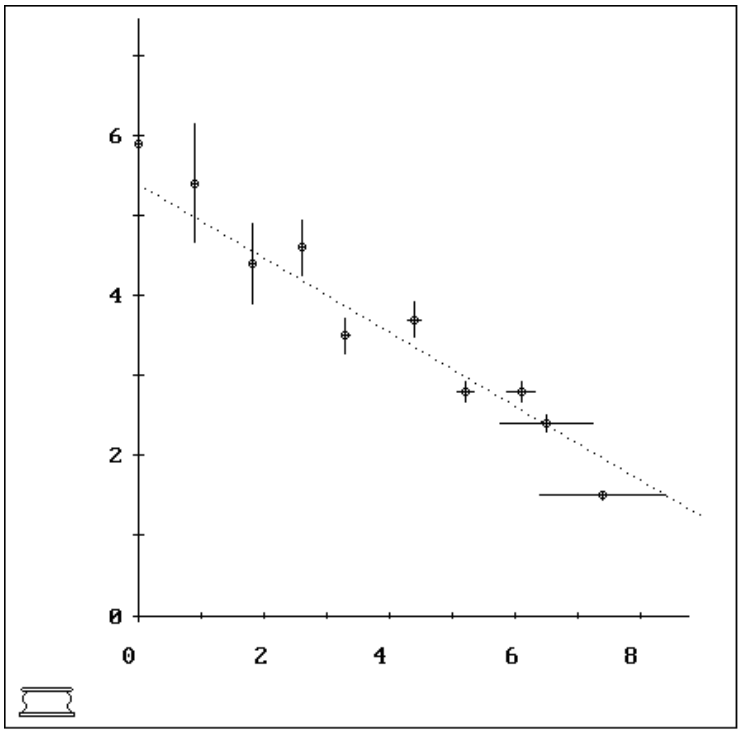



```
sigxy

XOFF =  0
  DX =  1
YOFF =  0
  DY =  1

 Y = A+B*X
(N = 10 )    IT# 3
    A =  5.394347
SIG(A) =  .293262
    B = -.462805
SIG(B) =  5.740726E-02
CHI^2/(N-2) =
          1.519665
```

The 3rd iteration plotted with s_x and s_y .




```
sigxy

XOFF =  0
  DX =  1
YOFF =  0
  DY =  1

 Y = A+B*X
(N = 10 )    IT# 3
    A =  5.394347
SIG(A) =  .293262
    B = -.462805
SIG(B) =  5.740726E-02
CHI^2/(N-2) =
          1.519665
```

SUMMARY

Weighted L.S.F. to y = a + bx

$$b = \frac{1}{\Delta}\left(\sum\frac{1}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2}\right)$$

$$a = \frac{\sum y_i/\sigma_i^2 - b\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$$

$$\sigma_a^2 \cong \frac{1}{\Delta}\sum x_i^2/\sigma_i^2 \qquad \text{and} \qquad \sigma_b^2 \cong \frac{1}{\Delta}\sum 1/\sigma_i^2$$

where $$\Delta = \sum 1/\sigma_i^2 \sum x_i^2/\sigma_i^2 - \left(\sum x_i/\sigma_i^2\right)^2$$

and $$\frac{\chi^2}{N-2} = \frac{1}{N-2}\sum\frac{1}{\sigma_i^2}(y_i - y_f)^2$$

Equal Weights i.e. , $\sigma_i = \sigma$

$$b = \frac{1}{\Delta'}(N\sum x_i y_i - \sum x_i \sum y_i)$$

$$a = \frac{1}{N}(\sum y_i - b\sum x_i)$$

$$\sigma_a^2 \cong \frac{\sigma^2}{\Delta'}\sum x_i^2 \qquad \text{and} \qquad \sigma_b^2 \cong \frac{N\sigma^2}{\Delta'}$$

where $$\Delta' = N\sum x_i^2 - (\sum x_i)^2$$

Unweighted L.S.F.

The equations are the same as the equal weights case with `s` approximated by the variance

$$\sigma^2 \cong \frac{1}{N-2}(\sum y_i^2 - a\sum y_i - b\sum x_i y_i)$$

Note: In a program, an Unweighted L.S.F. can be obtained using the Weighted L.S.F. equations by setting `s` = 1. The sigmas on the slope and intercept thus generated must be multiplied by the variance to obtain the proper sigmas for the Unweighted case.

Appendix

The origin of the effective variance is suggested by looking at the equation for the deviation

$$d_i = \Delta y_i = y_i - a - bx_i$$

Using Eq. (8) page 1-5 gives

$$\sigma_d^2 = \sigma_y^2 + b^2\sigma_x^2$$

J.Orear* has shown that the effective variance technique is more generally stated as

$$\sigma_i^2 = \sigma_y^2 + \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2$$

and is exact in the limit that $\partial f / \partial x$ is constant over the region of σ_x. This means it is always exact for linear fits. Also the iterative effective variance procedure works for any functional form of $f(x)$.

Linearized Fits

$f(x) = z = me^{bx}$ or $\ln z = \ln m + bx$

$$\frac{\partial f(x)}{\partial x} = bme^{bx} = bz$$

$$\sigma_i^2 = \sigma_z^2 + (\partial f / \partial x)^2 \sigma_x^2 = \sigma_z^2 + (bz)^2 \sigma_x^2$$

$$\boxed{\left(\frac{\sigma_i}{z}\right)^2 = \left(\frac{\sigma_z}{z}\right)^2 + b^2\sigma_x^2}$$

Data is entered in Equations (3), (4), (9), and (15) with $x_i = x_j$, $y_i = \ln z_j$, etc. (See page 4-11).

* Least Squares when both variables have uncertainties, CLNS-81/496 June, Cornell University, Ithaca, N.Y.

Chi-Squared Minimization

(In collaboration with - Anastasios Vayonakis)

The limitations of iterating vertical minimization demonstrated by Pearson's data in Chapter 4 will be dealt with in this chapter. The most general statement of χ^2, for a point with uncertainties in both x and y , is

$$\chi^2 = \sum_{i=1}^{N} \frac{(y-y_i)^2}{\sigma y_i^{\,2}} + \frac{(x-x_i)^2}{\sigma x_i^{\,2}}$$

where x, y defines a point on the fit line.

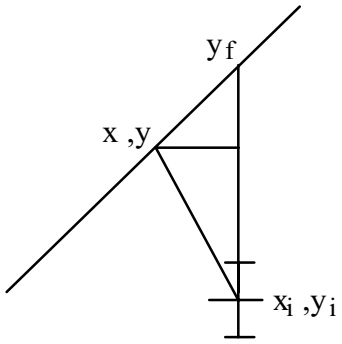


Figure 1

b = slope
b = $(y_f - y)/(x_i - x)$

$y - y_i = (y_f - y_i) - (y_f - y)$
$= (y_f - y_i) - b(x_i - x)$

What we will show next is that this non-vertical minimization along a line from the point x_i, y_i with uncertainties σx and σy to x, y on the fit line, is equivalent to the more familiar vertical minimization with an uncertainty in the form of the effective variance $\sigma_i^2 = \sigma y_i^2 + b^2 \sigma x_i^2$.

For each point (x_i , $y_i \pm \sigma y_i$)

$$P_i(y) = \frac{1}{\sqrt{2\pi}\,\sigma y_i} \exp\left(-\frac{(y-y_i)^2}{2\sigma y_i^2} \right)$$

If we have uncertainties in both x and y : ($x_i \pm \sigma x_i$, $y_i \pm \sigma y_i$), then the probability distribution of point i is:

$$P_i(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma y_i \sqrt{2\pi}\,\sigma x_i} \exp\left(-\frac{(y-y_i)^2}{2\sigma y_i^2} - \frac{(x-x_i)^2}{2\sigma x_i^2} \right)$$

Using the expression for $y - y_i$ in Figure 1 and integrating over all x gives:

$$P_i(y) = \int_{-\infty}^{\infty} P_i(x,y)\,dx$$

$$P_i(y) = \frac{1}{2\pi\sigma y_i \sigma x_i} \exp\left(-\frac{\Delta y_i^{\,2}}{2\sigma y_i^{\,2}} \right) \int_{-\infty}^{\infty} \exp\left(-\frac{b^2 (x - x_i)^2}{2\sigma y_i^{\,2}} - \frac{b\,\Delta y_i (x - x_i)}{\sigma y_i^{\,2}} - \frac{(x - x_i)^2}{2\sigma x_i^{\,2}} \right) dx$$

where $\Delta y_i = y_f - y_i$. Now change variables letting $z = x - x_i$

$$P_i(y) = \frac{1}{2\pi\sigma y_i \sigma x_i} \exp\left(-\frac{\Delta y_i^{\,2}}{2\sigma y_i^{\,2}} \right) \int_{-\infty}^{\infty} \exp\left(-\left\{ \frac{b^2}{2\sigma y_i^{\,2}} + \frac{1}{2\sigma x_i^{\,2}} \right\} z^2 - \frac{b\,\Delta y_i}{\sigma y_i^{\,2}} z \right) dz$$

The above integral is of the form

$$\int_{-\infty}^{\infty} \exp\left(-\alpha^2 z^2 \pm \beta z \right) dz = \frac{\sqrt{\pi}}{|\alpha|} \exp\left(\frac{\beta^2}{4\alpha^2} \right)$$

giving

$$P_i(y) = \frac{1}{\sqrt{2\pi}\ \sqrt{\sigma y_i^{\,2} + b^2 \sigma x_i^{\,2}}} \exp\left(\frac{\Delta y_i^{\,2}}{2\left(\sigma y_i^{\,2} + b^2 \sigma x_i^{\,2} \right)} \right)$$

Thus vertical minimization modified by the effective variance is equivalent to the more general non-vertical minimization.

We are thus faced with minimizing χ^2 in the form

$$\chi^2 = \sum_i^N \frac{\left(a + b x_i - y_i \right)^2}{\sigma y_i^{\,2} + b^2 \sigma x_i^{\,2}} \tag{1}$$

The b^2 term in the denominator prevents an analytic solution from being found, and any iterative solution will lead to the results of Chapter 4. The best fit results will be obtained by putting χ^2 into a minimization program, such as the one in Mathematica, and minimizing the entire function.

The b^2 term in the denominator is the slope which in more general terms for a function $y = F(x)$ is $y' = \partial F(x) / \partial x$ giving the more general expression for χ^2 as

$$\chi^2 = \sum_i^N \frac{\left(F(x) - y_i \right)^2}{\sigma y_i^{\,2} + \left(\frac{\partial F}{\partial x} \right)^2 \sigma x_i^{\,2}} \tag{2}$$

(See Appendix Chapt. 4)

Now let's estimate the uncertainties, σ_a & σ_b, related to the best fit values a^* & b^* that minimize χ^2. In the region of a^* the parameter a, has a distribution that is close to being Gaussian. In fact the larger the data set, the closer this distribution is to that of a Gaussian probability distribution. (This follows from the Central Limit Theorem).

Thus $P(a)$ is approximated by

$$P(a) = A\exp\left(-\frac{(a-a^*)^2}{2\sigma_a^2}\right)$$

At $a = a^*$, $P(a^*) = A$

At $a = a^*+\sigma_a$ $\qquad P(a^*+\sigma_a) = P(a^*)\exp(-1/2)$ (3)

In χ^2 minimization, a & b are treated as variables when seaching for a^* and b^*. With this in mind (and using Eq. (2)) let's define $P(a,b)$ as

$$P(a,b) = A\exp\left(-\sum_i \frac{(y_i - y_f)^2}{2\sigma_i^2}\right) = A\exp\left(-\tfrac{1}{2}\chi^2(a,b)\right)$$

At $a = a^*$, $b = b^*$ $\qquad A = \dfrac{P(a^*,b^*)}{\exp\left(-\frac{1}{2}\chi^2(a^*,b^*)\right)}$

At $a = a^*+\sigma_a$, b

$$P(a^*+\sigma_a,b) = P(a^*,b^*)\exp\left(-\tfrac{1}{2}\left(\chi^2(a^*+\sigma_a,b) - \chi^2(a^*,b^*)\right)\right)$$

To give a result similar to equation (3) requires that

$$\underline{\underline{\chi^2(a^*+\sigma_a,b) - \chi^2(a^*,b^*) = 1}} \qquad (4)$$

Thus the uncertainty σ_a, is that change in a which increases the value of χ^2 by one unit larger than the minimum value $\chi^2(a^*,b^*)$

$$\chi^2(a = a^* + \sigma_a,\ b) = \chi^2(a^*,b^*) + 1$$

Similarly at $b = b^*+\sigma_b$

$$\chi^2(a,\ b = b^*+\sigma_b) = \chi^2(a^*,b^*) + 1$$

A subtle but important point is that each time a is varied from a^*, while searching for the value of a that increases $\chi^2(a^*,b^*)$ by one unit, the other parameters must be reminimized locally, at this new value of a, before χ^2 is calculated.

The above results are summarized in Figure 2.

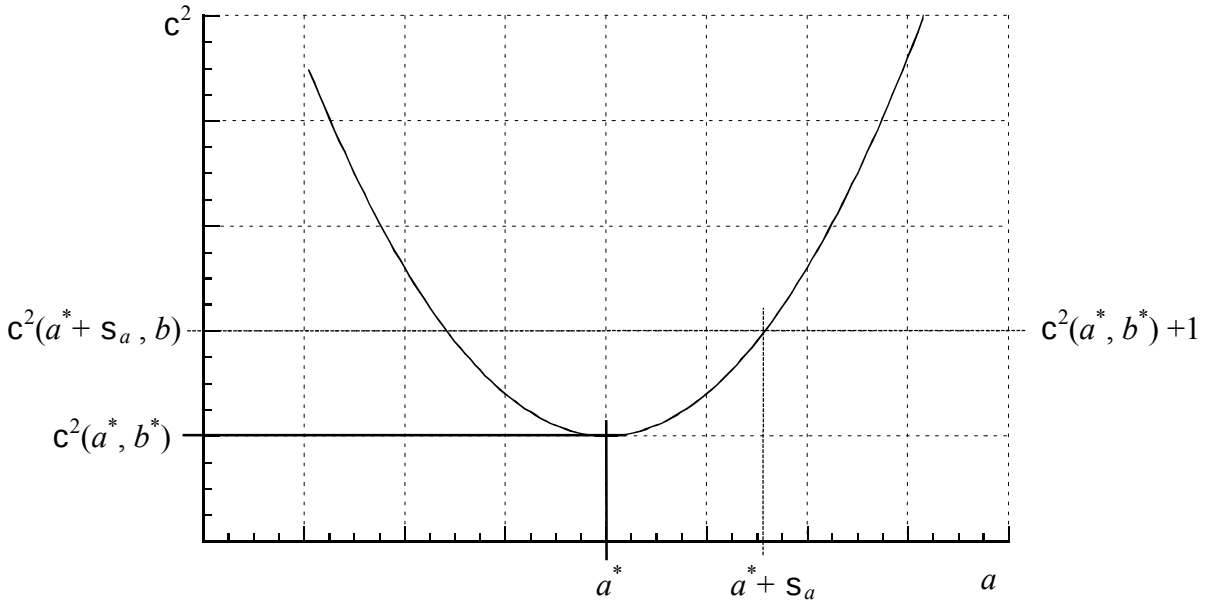


Figure 2

Finally the above technique gives a value for σ_a for any value of χ^2. In order to check the validity of σ_a it is necessary to also calculate the Reduced χ^2 [$\chi^2/(N\text{-}2)$]. If the Reduced Chi-Squared varies significantly from one, a more realistic value for σ_a may be $\sigma_a \sqrt{\chi^2/(N-2)}$.

The ideas presented in these five chapters are implemented in a Mathematica 3 notebook "CurveFit" available on the Caltech WEB site at

www.pma.caltech.edu/~dons/ph3-7/

A Linux version of these ideas is available in the program "Glove" available at

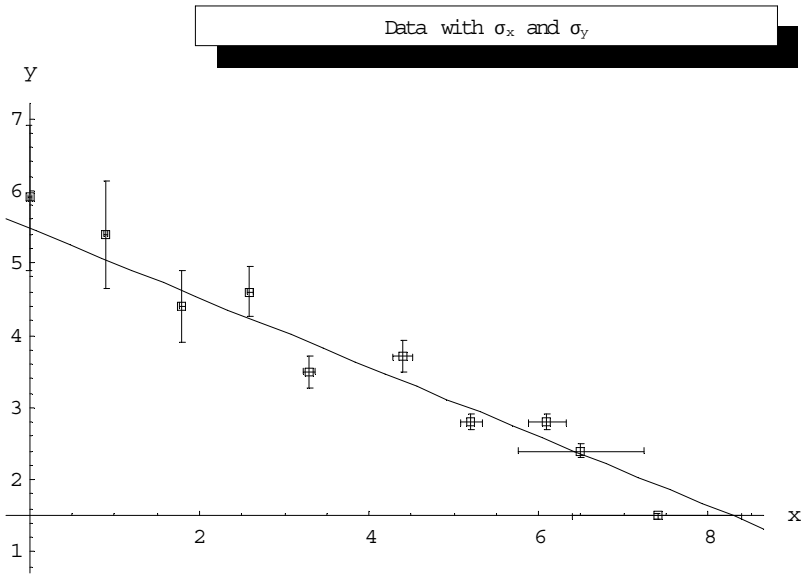
www.cco.caltech.edu/~glenn/glove/

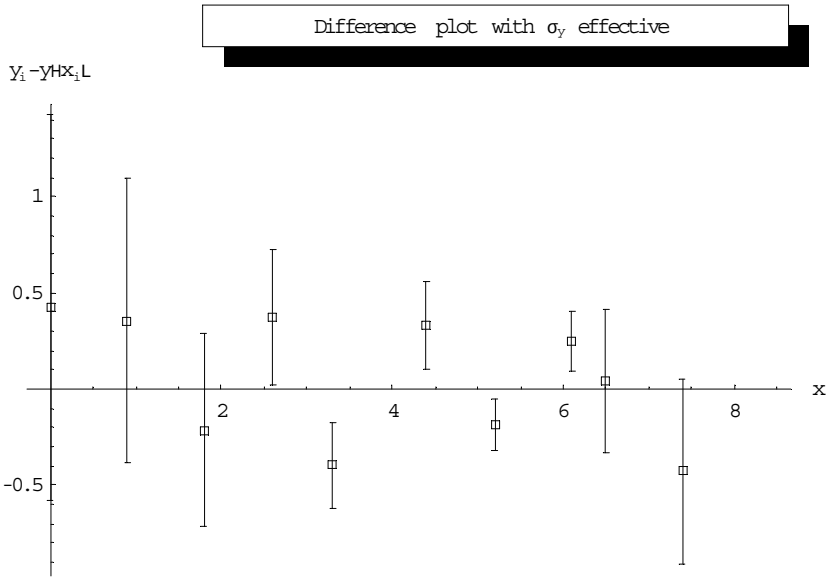
The data on page 4-18 when fitted with CurveFit yeilds invariant results given by

$$y(x) = a + bx$$

$$a = 5.47 \pm 0.29 \quad b = -0.479 \pm 0.059$$

$$c^2/(n-2) = 1.508$$





The Argument for Always Including Uncertainties in the Independent Variable

Let's look at the sensitivity of a particular fit to uncertainties in x relative to the usual uncertainties in y. Let's start with the linear case $\boxed{y = a + bx}$. Solving for x and differentiating gives

$$\sigma_x = \frac{\sigma_y}{b} . \tag{5}$$

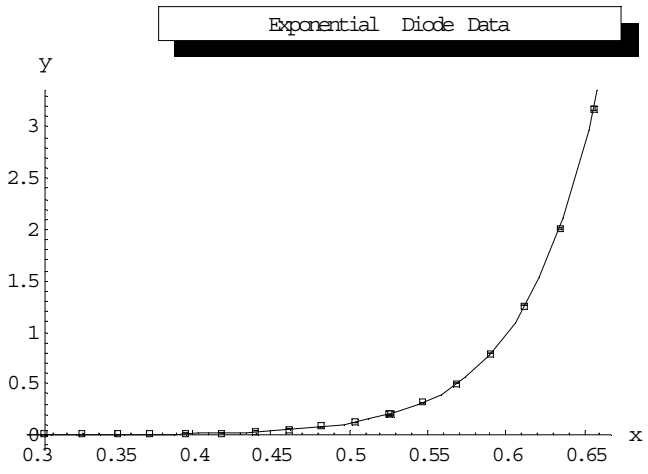
This implies that if b is large the equivalent s_x, which has a similar effect on the fit as s_y, is some small fraction of s_y. If b is small, but greater than one, the equivalent s_x is similar to s_y. If b is less than one the fit is dominated by s_y, becoming less and less sensitive to s_x as b approaches zero. Similar conclusions are obtained from the effective variance equation

$$\sigma'^2 = \sigma_y^2 + b^2\sigma_x^2$$

Next let's look at exponential data described by $\boxed{y = a\,e^{bx}}$. Taking the log, solving for x, and differentiating, gives

$$\sigma_x = \frac{1}{b}\frac{\sigma_y}{y} . \tag{6}$$

The conclusions regarding the effectiveness of s_x, for various values of b, are similar to the conclusions in the linear case, with the additional sensitivity due to the $1/y$ factor. Thus for this exponential example there is sensitivity related to y; the larger the y value the greater the sensitivity to s_x. This is clearly illustrated in the linear plot of some diode data below.



In the more general case of
$y = a\,e^{bx} + c$
the analysis gives

$$\ln y = \ln a + bx + \ln\left(1 + c/a e^{bx}\right)$$

$$bx = \ln y - \ln a - \ln\left(1 + c/y - c\right) .$$

Giving $\sigma_x = \frac{1}{b}\frac{\sigma_y}{y}\left[1 + \frac{c}{y-c}\right]$.

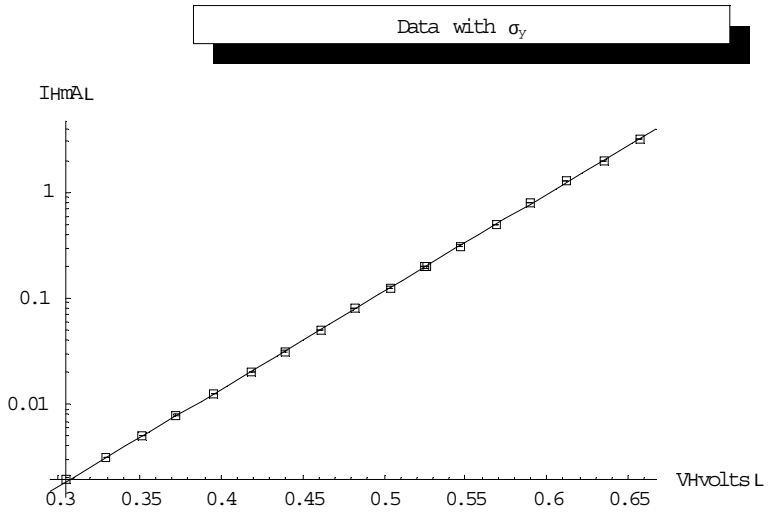
If $y >> c$ the result reduces to Eq.(6) and when y is comparable to c the slope b is approaching 0 where the fit is insensitive to s_x

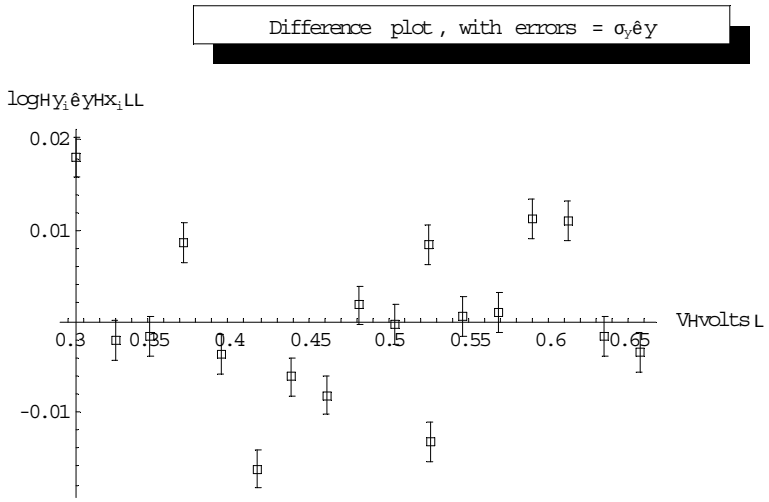
For log-log plots where $\boxed{y = a\,x^b}$ similar analysis gives

$$\frac{\sigma_x}{x} = \frac{1}{b}\frac{\sigma_y}{y} \tag{7}$$

Again, for a similar effect on the analysis, the relative σ_x/x is $1/b$ smaller than the % error in y that has the same effect on the results.

For illustration let's look at some silicone diode data where $I = I_o e^{bV}$. Here $s_I = 0.5\% I$.



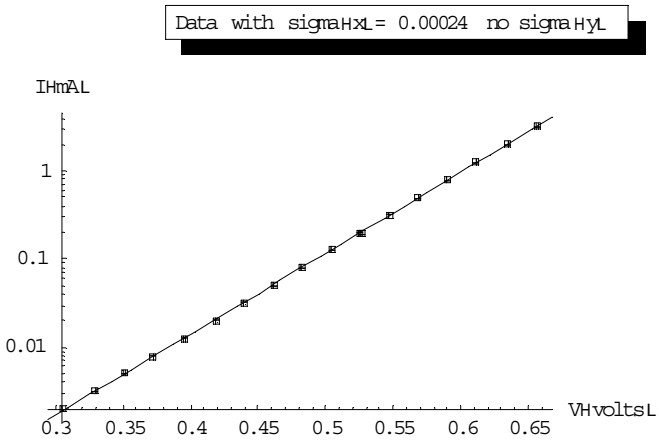


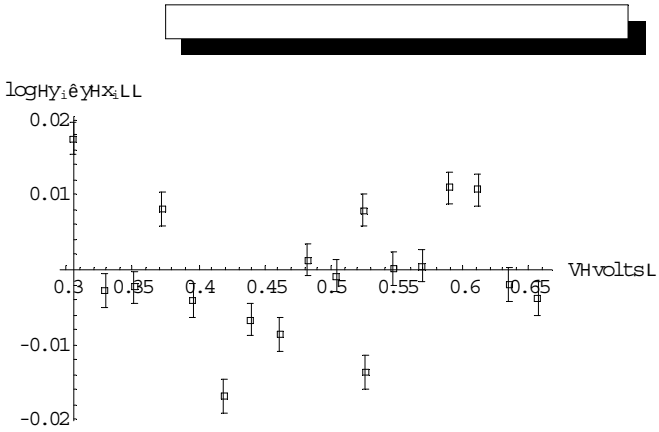
n= 18

yHxL = a expHb xL

a=	b=		
2.94365×10^{-6}	21.1661		
σ_a=	σ_b=	χ^2êHn−2L=	
1.6691×10^{-8}	0.0114239	17.2541	

The results imply that an equivalent s_x from Eq.(6) would be $s_x = 0.005/21.1 = 0.000237$





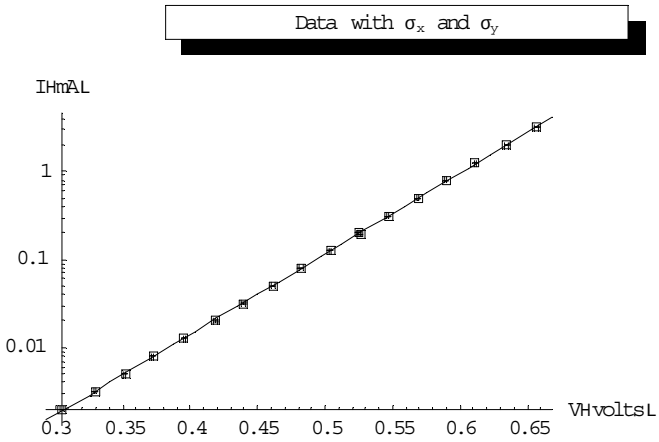
n= 18

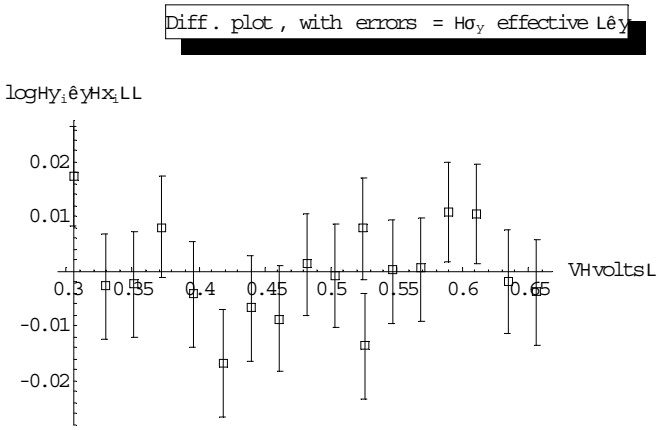
yHxL = a expHb xL

a=	b=	
2.95028×10^{-6}	21.1638	
σ_a=	σ_b=	χ^2êHn-2L=
1.65304×10^{-8}	0.0112793	17.078

From the similar appearance of the difference plots and the similar values of $c^2/(n-2)$ it is clear that Eq.(6) has produced an equivalent s_x , the value of which is approximately a factor of 10 smaller than s_y .

The actual digital voltage measurements had a s_x = .001 volts. When this value of s_x is used in the fit, the results account for the scatter in the data giving $c^2/(n-2)$ Ü 1.





n= 18

yHxL = a expHb xL

a=	b=	
2.94992×10^{-6}	21.164	
σ_a=	σ_b=	χ^2êHn-2L=
7.20765×10^{-8}	0.0488268	0.915992

It is clear from both the Difference plot and the value of $c^2/(n-2)$ that including the part in a thousand uncertainty in the voltage measurement accounts for all the scatter in the data

The above arguments indicate that the sensitivity to errors in x in a given fit may depend on several factors, type of fit, values of fit parameters such as b, etc. The values of fit parameters are often not know when starting the fit procedure, so in most cases their impact on the sensitivity of the fit to $\mathtt{s}_x$ is not realized. I have periodically been surprised by the sensitivity to errors in x, of data for which it was not intuitvely obvious that the data should be that sensitive.

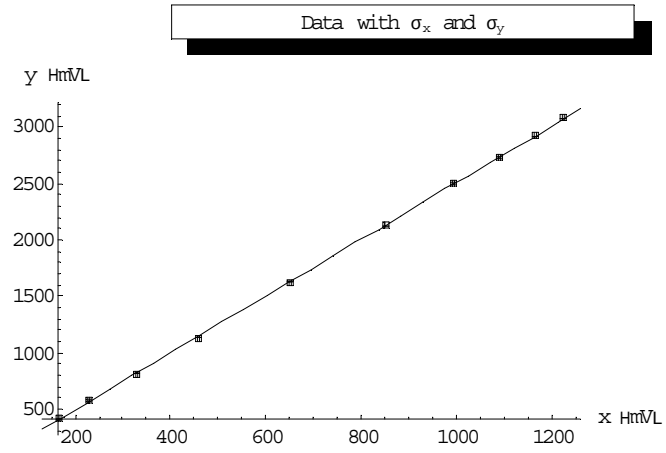
With so much data taken digitally today, the variables all have similar uncertainties associated with them. Doing analysis using only $\mathtt{s}_y$,which the fit may be least sensitive to, seems incomplete. If the analysis can only deal with uncertainties in one variable you may have to rewrite the fitting equation to make that variable the dependent variable. Thus $I = I_o \mathrm{e}^{bV}$ becomes $V = A + 1/b \ln I$ if the only uncertainties available are $\mathtt{s}_V$.

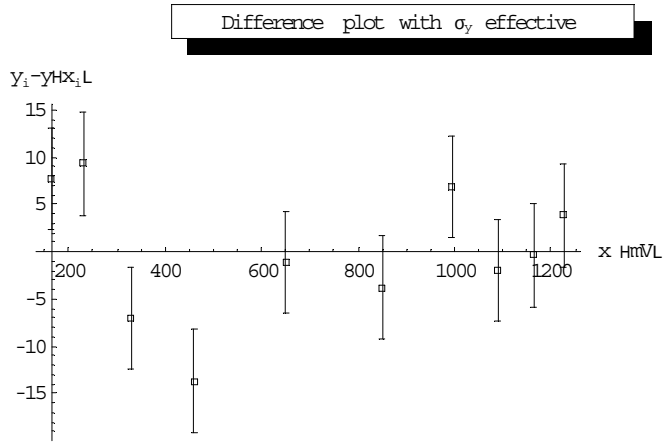
Finally the role of $\mathtt{s}_x$ in helping to account for the scatter in the data is important in evaluating whether the analysis has accounted for all the effects impacting the data.

An example of digital data

point	x	y	σ_y	σ_x	z
1	165.	415.	2.	2.	1
2	230.	580.	2.	2.	1
3	330.	815.	2.	2.	1
4	460.	1135.	2.	2.	1
5	650.	1625.	2.	2.	1
6	850.	2125.	2.	2.	1
7	995.	2500.	2.	2.	1
8	1090.	2730.	2.	2.	1
9	1165.	2920.	2.	2.	1
10	1225.	3075.	2.	2.	1

Both the x and y data were obtained with 5 Volt, 10 bit ADC's. The minimum step is $5V/2^{10}$ Ü $5Õ10^{-3}$ V Ü 5 mV. The uncertainty should be some fraction of 5 mV say 2 mV. This uncertainty is the same for both variables.

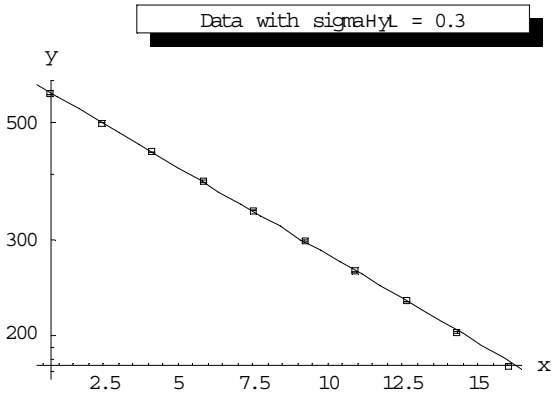


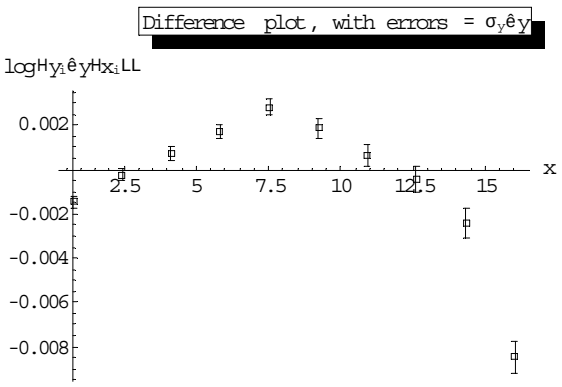


yHxL = a + b x

a=	b=	
−7.33973	2.51307	
σ_a=	σ_b=	χ^2êHn−2L=
3.72174	0.00457706	1.97535

BEWARE the ADDITIVE CONSTANT



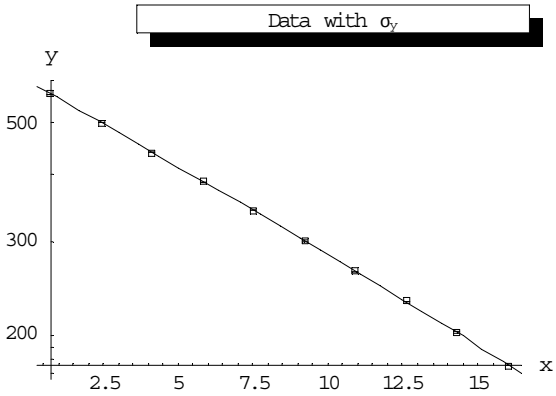


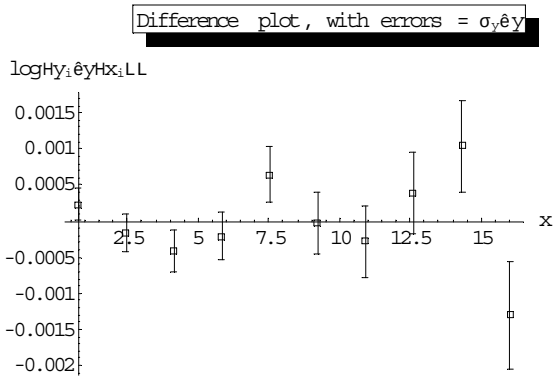
$$yHxL = a\ expHb\ xL$$

a=	b=		
601.128	−0.0757644		
σ_a=	σ_b=	χ^2êHn−2L=	
0.248686	0.0000624423	36.8387	

The usual exponential fit seen above often leads to apparent second order effects, as seen in the difference plot. These apparent 2nd order effects may lead to extensive discussions of possible 2nd order physical effects. Many times these apparent effects seen in the difference plot are due to an unexpected systematic constant in the data which becomes clear when the data is fitted with

$$yHxL = a\ expHb\ xL + c$$





a=	b=	c=	
627.606	−0.0696509	−29.9669	
σ_a=	σ_b=	σ_c=	χ^2êHn−3L=
1.75018	0.000368259	1.94582	1.86118

The added constant *c*, may be an unexpected offset or an artifact of some physical effect. Further examination of the experimental equipment and the data may be required.

Final Examples

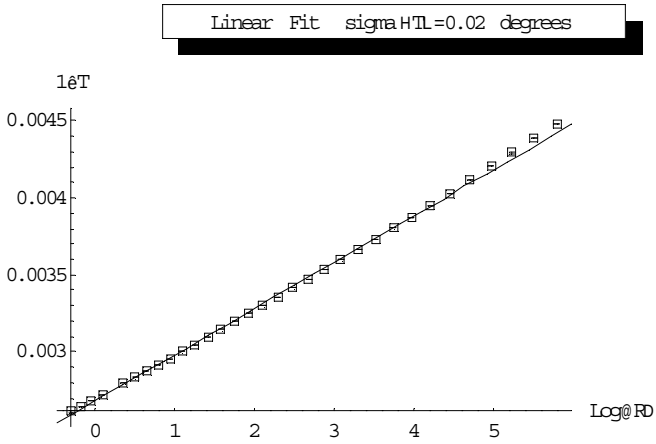
As one example of the ideas in the Primer let's look at some typical thermistor data. Thermistors are usually made of a semiconductor material with a band gap E_g. The response is described by the equation $R \propto R_o e^{E_g/2kT}$ (R is resistance, T is temp. in ^oK) (8)

Which can be rewritten as $\ln[R] = \ln[R_o] + \frac{E_g}{2k}\frac{1}{T}$ or $\frac{1}{T} = a + b\ln[R]$ (9)

The band gap E_g, has a small variation with temperature which may explain the more complicated response of a typical thermistor. Correction terms can be generated by doing a polynomial expansion in ln[R] giving

$$\frac{1}{T} = a + b\ln[R] + c(\ln[R])^2 + d(\ln[R])^3 + \cdots \quad (10)$$

Let's start with a linear fit to some typical thermistor data (see page 5-19) $\frac{1}{T} = a + b\ln[R]$

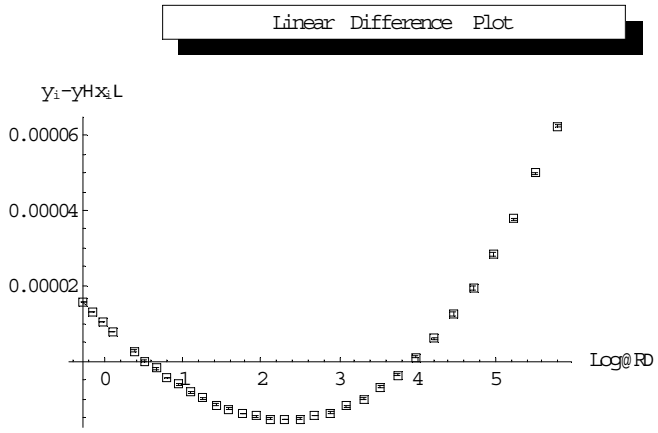


n= 32

yHxL = a + b x

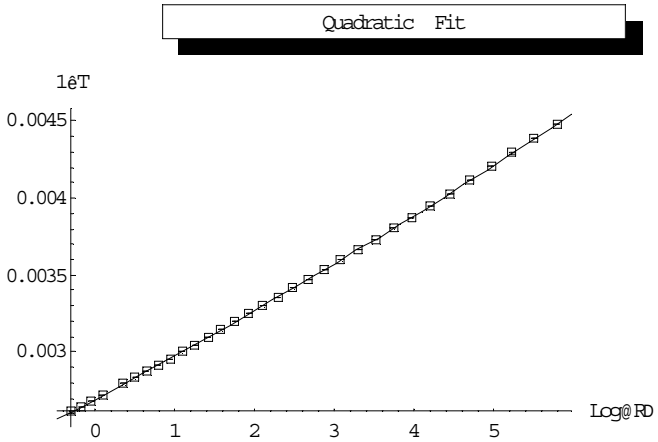
a=	b=	
0.00267867	0.000300741	
σ_a=	σ_b=	χ^2êHn-2L=
5.00274×10^{-8}	2.43186×10^{-8}	4727.51

It is clear from the appearance of the plot and the reduced c^2, that there is more than just linear information in this data set. Let's look at the difference plot.



There is an apparent quadratic effect in the data. Let's try fitting the data with

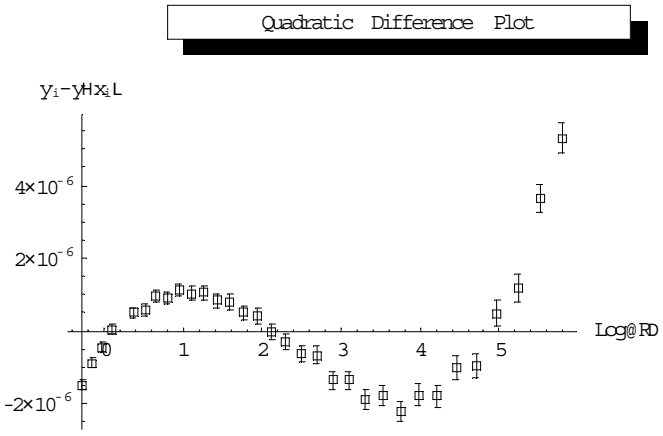
$$\frac{1}{T} = a + b\ln[R] + c(\ln[R])^2$$



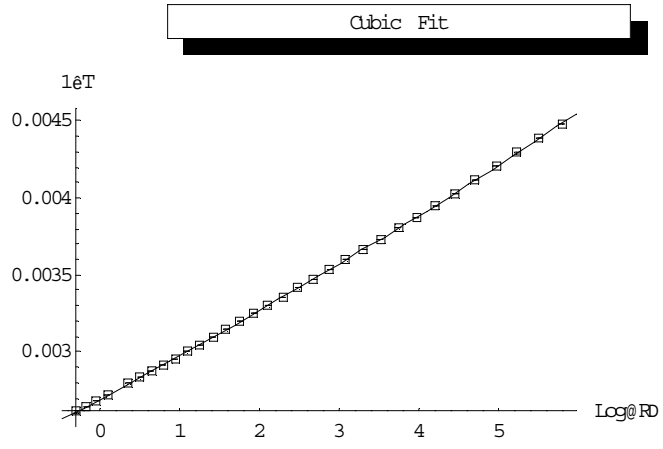
$y(x) = a + b\,x + c\,x^2$

a=	b=	c=	
0.00268886	0.00027717	5.4602×10^{-6}	
σ_a=	σ_b=	σ_c=	$\chi^2/(n-3)$=
5.69126×10^{-8}	6.73538×10^{-8}	1.45498×10^{-8}	34.213

Hard to see any higher order effects in this plot. $c^2/(n-3)$ is hinting at something. Let's look at the difference plot.

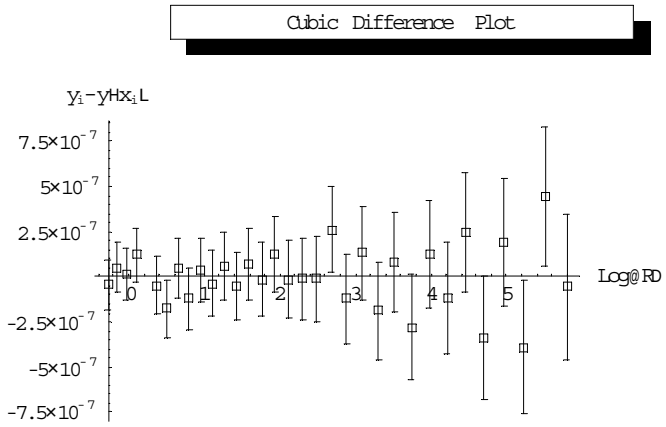


A rather pronounced residual is still in the data. Let's try a cubic fit by adding the $d(\ln[R])^3$ term.



$$yHxL = a + b\,x + c\,x^2 + d\,x^3$$

a=	b=	
0.00268848	0.000280392	
σ_a=	σ_b=	
5.81934×10^{-8}	1.22968×10^{-7}	
c=	d=	
3.41795×10^{-6}	2.8593×10^{-7}	
σ_c=	σ_d=	χ^2êHn-4L=
6.681×10^{-8}	9.12958×10^{-9}	0.403506



Finally, there is no obvious residual information left in the data set. The uncertainty in the temperature of 0.02 degrees is a little too large, as it is not quite consistent with the scatter in the data. As a result χ^2/(n-4) is less than one and a better estimate of the uncertainties s_a , s_b , etc. would be $\sigma'_a = \sigma_a \sqrt{\chi^2/(n-4)} = 5.82\times10^{-8}\sqrt{0.403} = 3.69\times10^{-8}$ etc.

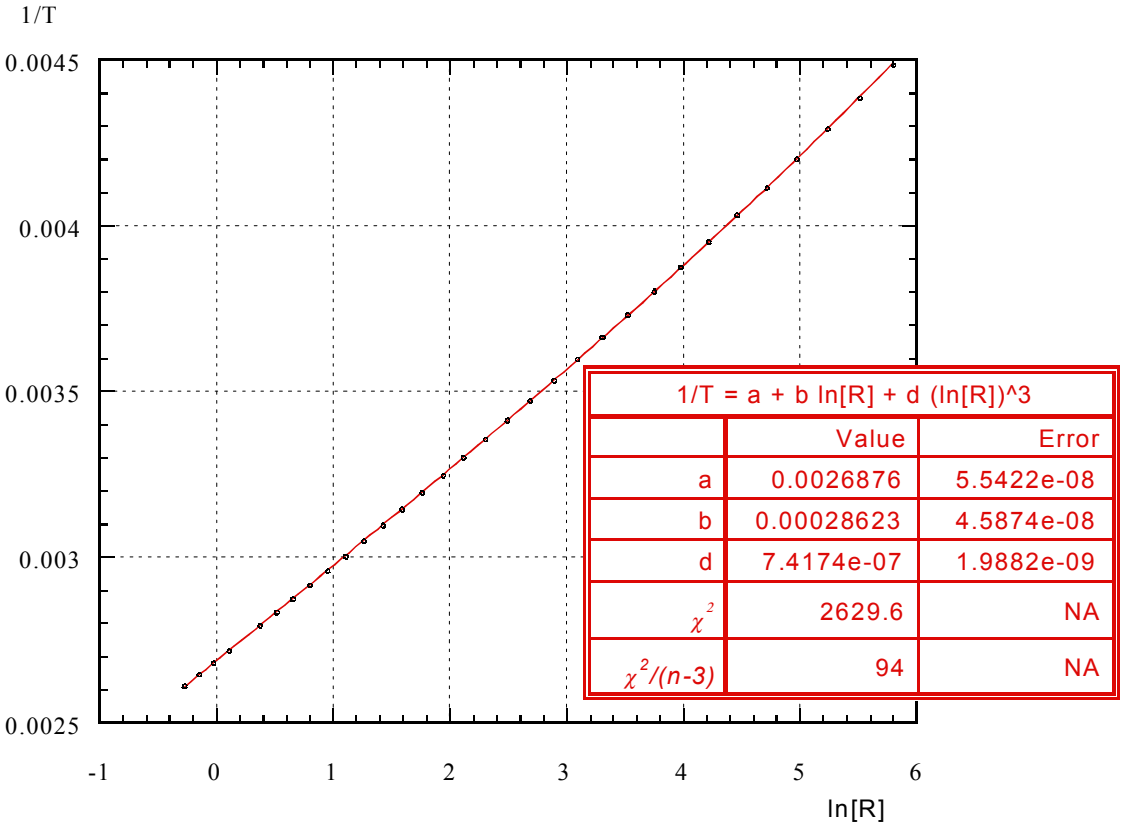
In 1968 Steinhart and Hart, using the tools available, concluded that their thermistor data was best fit by leaving out the quadratic term giving the Steinhart-Hart Equation[2]

$$\frac{1}{T} = a + b\ln[R] + c(\ln[R])^3$$

A fit to the above thermistor data using the Steinhart-Hart Equation is given on the next page. You make your own judgement as to whether this is an appropriate fit to the data.

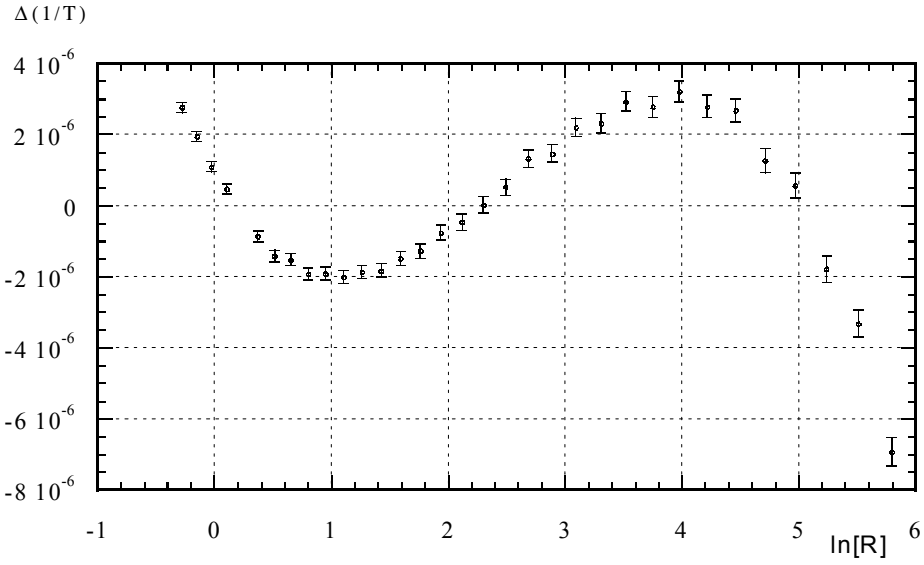
[2] Deep Sea Research, 1968, Vol. 15, pp 497 to 501. Pergamon Press Printed in Great Britain.

Steinhart-Hart Fit to the Thermistor Data



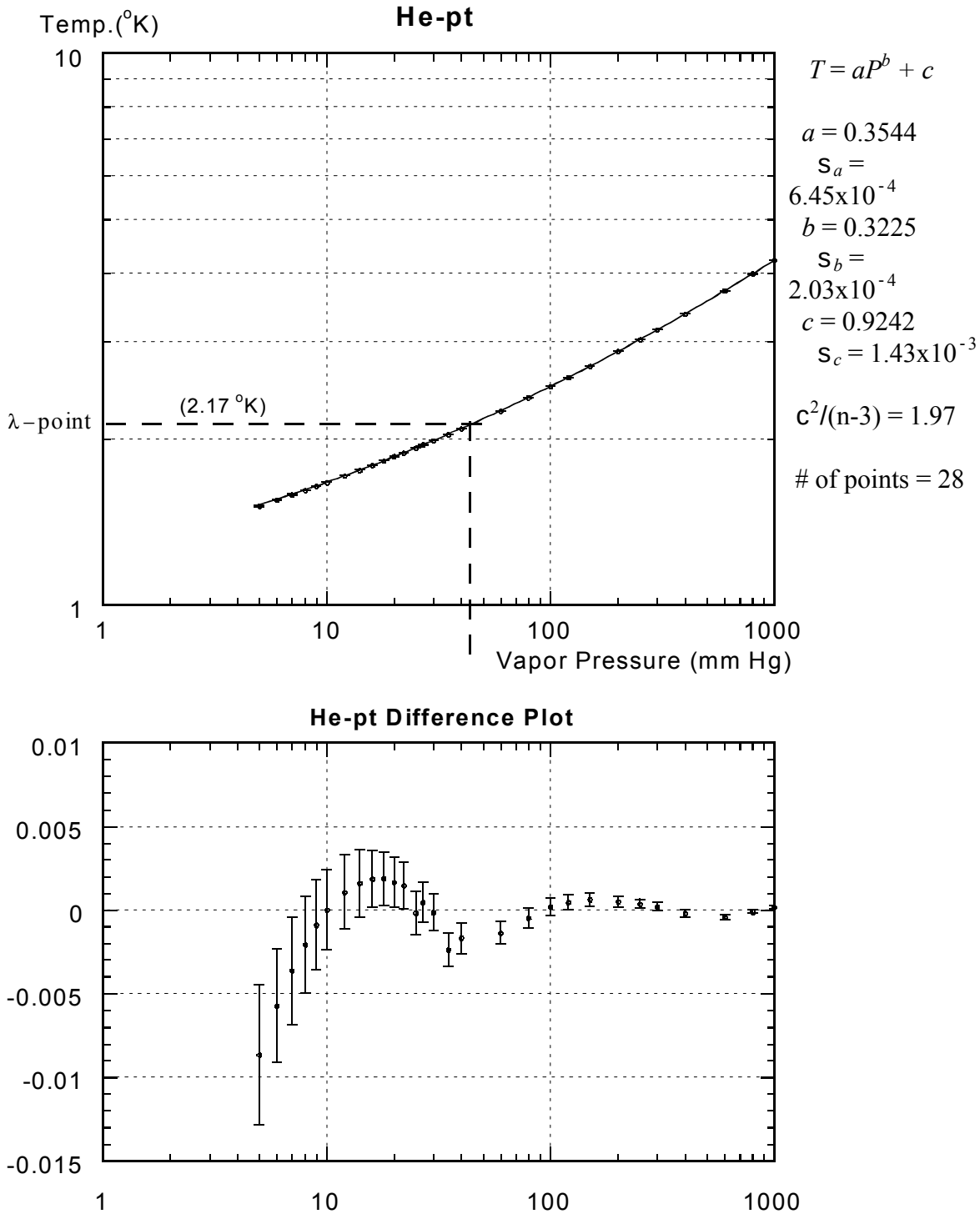
1/T = a + b ln[R] + d (ln[R])^3		
	Value	Error
a	0.0026876	5.5422e-08
b	0.00028623	4.5874e-08
d	7.4174e-07	1.9882e-09
χ^2	2629.6	NA
$\chi^2/(n-3)$	94	NA

S-H Difference Plot



Second Example

The power law relationship between pressure and temperature for liquid He is seen in the next plot. The plot of the data and the fit parameters show no hint of the complexity of the



physical situation clearly exhibited in the difference plot.

Thermistor data plotted on pages 5-12 through 5-16

point	x (Resist.)	y (Temp) in C	σ_y	σ_x
1	0.7579	110.	0.02	0
2	0.8575	105.	0.02	0
3	0.9735	100.	0.02	0
4	1.108	95.	0.02	0
5	1.451	85.	0.02	0
6	1.669	80.	0.02	0
7	1.924	75.	0.02	0
8	2.229	70.	0.02	0
9	2.589	65.	0.02	0
10	3.021	60.	0.02	0
11	3.537	55.	0.02	0
12	4.161	50.	0.02	0
13	4.912	45.	0.02	0
14	5.828	40.	0.02	0
15	6.941	35.	0.02	0
16	8.313	30.	0.02	0
17	10.	25.	0.02	0
18	12.09	20.	0.02	0
19	14.68	15.	0.02	0
20	17.96	10.	0.02	0
21	22.05	5.	0.02	0
22	27.28	0	0.02	0
23	33.89	-5.	0.02	0
24	42.45	-10.	0.02	0
25	53.39	-15.	0.02	0
26	67.74	-20.	0.02	0
27	86.39	-25.	0.02	0
28	111.3	-30.	0.02	0
29	144.	-35.	0.02	0
30	188.4	-40.	0.02	0
31	247.5	-45.	0.02	0
32	329.2	-50.	0.02	0

APPENDIX
LEAST SQUARE FITS - MATRIX METHOD

LINEAR FIT:

Let's find the fit to $\boxed{y = B_1 + B_2 x}$ (1)

The probability is proportional to

$$P \propto \prod_i \exp\left[(B_1 + B_2 x_i - y_i)^2 / \sigma_i^2\right] \tag{2}$$

Taking the log of Eq.(2)

$$W = \ln P \propto \sum_i (B_1 + B_2 x_i - y_i)^2 / \sigma_i^2 \tag{3}$$

Now let's optimize W

$$\frac{\partial W}{\partial B_1} = 0 \qquad \text{and} \qquad \frac{\partial W}{\partial B_2} = 0 \tag{4}$$

From Eq.(3) and (4) we obtain

$$\sum_i \frac{1}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i}{\sigma_i^2} \times B_2 = \sum_i \frac{y_i}{\sigma_i^2} \tag{5}$$

and

$$\sum_i \frac{x_i}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i^2}{\sigma_i^2} \times B_2 = \sum_i \frac{x_i y_i}{\sigma_i^2} \tag{6}$$

Equations (5) and (6) can be written in matrix form as

$$\mathbf{AB} = \mathbf{V} \tag{7}$$

where

$$\mathbf{A} = \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} \\ \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{x_i^2}{\sigma_i^2} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \qquad \mathbf{V} = \begin{bmatrix} \sum_i \frac{y_i}{\sigma_i^2} \\ \sum_i \frac{x_i y_i}{\sigma_i^2} \end{bmatrix}$$

Now let's multiply Eq.(7) by the inverse of the matrix **A**

$$\mathbf{A}^- \mathbf{AB} = \mathbf{B} = \mathbf{A}^- \mathbf{V} \tag{8}$$

So all we have to do to find the fit parameters B_1 and B_2, is to invert the **A** matrix. The $\mathbf{A}^-$ matrix is also known as the error matrix with $\sigma_{B_1}^2 = \mathrm{A}_{11}^-$, and $\sigma_{B_2}^2 = \mathrm{A}_{22}^-$, also with A_{12}^- and A_{21}^- the covariant terms (s_{ab}) [See Eq.(9) page 4-3 and page 4-13]

QUADRATIC FIT:

Let's find the fit to $\boxed{y = B_1 + B_2 x + B_3 x^2}$ (9)

The probability is proportional to

$$P \propto \prod_i \exp\left[(B_1 + B_2 x_i + B_3 x_i^2 - y_i)^2 / \sigma_i^2\right] \tag{10}$$

Taking the log of Eq.(10)

$$W = \ln P \propto \sum_i (B_1 + B_2 x_i + B_3 x_i^2 - y_i)^2 / \sigma_i^2 \tag{11}$$

Optimizing W

$$\frac{\partial W}{\partial B_1} = 0 \qquad \frac{\partial W}{\partial B_2} = 0 \quad \text{and} \quad \frac{\partial W}{\partial B_3} = 0 \tag{12}$$

From Eqs.(11) and (12) we obtain

$$\sum_i \frac{1}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i}{\sigma_i^2} \times B_2 + \sum_i \frac{x_i^2}{\sigma_i^2} \times B_3 = \sum_i \frac{y_i}{\sigma_i^2} \tag{13}$$

$$\sum_i \frac{x_i}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i^2}{\sigma_i^2} \times B_2 + \sum_i \frac{x_i^3}{\sigma_i^2} \times B_3 = \sum_i \frac{x_i y_i}{\sigma_i^2} \tag{14}$$

$$\sum_i \frac{x_i^2}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i^3}{\sigma_i^2} \times B_2 + \sum_i \frac{x_i^4}{\sigma_i^2} \times B_3 = \sum_i \frac{x_i^2 y_i}{\sigma_i^2} \tag{15}$$

Again Eqs.(13) - (15) can be written in matrix form as

$$\mathbf{AB} = \mathbf{V} \tag{16}$$

where

$$\mathbf{A} = \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{x_i^2}{\sigma_i^2} \\ \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{x_i^2}{\sigma_i^2} & \sum_i \frac{x_i^3}{\sigma_i^2} \\ \sum_i \frac{x_i^2}{\sigma_i^2} & \sum_i \frac{x_i^3}{\sigma_i^2} & \sum_i \frac{x_i^4}{\sigma_i^2} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \quad \text{and} \quad \mathbf{V} = \begin{bmatrix} \sum_i \frac{y_i}{\sigma_i^2} \\ \sum_i \frac{x_i y_i}{\sigma_i^2} \\ \sum_i \frac{x_i^2 y_i}{\sigma_i^2} \end{bmatrix}$$

Now let's multiply Eq.(16) by the inverse of the matrix **A**

$$\mathbf{A}^{-}\mathbf{AB} = \mathbf{B} = \mathbf{A}^{-}\mathbf{V} \tag{17}$$

Once again all we have to do to find the fit parameters B_1 , B_2 and B_3 is to invert the **A** matrix. The $\mathbf{A}^{-}$ matrix is also the error matrix.

LINEAR FIT to TWO INDEPENDENT VARIABLES:

Let's find the fit to $\boxed{y = B_1 + B_2 x + B_3 z}$ (18)

where x and z are two independent variables.

The proability is proportional to

$$P \propto \prod_i \exp\left[(B_1 + B_2 x_i + B_3 z_i - y_i)^2 / \sigma_i^2\right] \tag{19}$$

Again taking the log of Eq.(19)

$$W = \ln P \propto \sum_i (B_1 + B_2 x_i + B_3 z_i - y_i)^2 / \sigma_i^2 \tag{20}$$

Optimizing W

$$\frac{\partial W}{\partial B_1} = 0 \qquad \frac{\partial W}{\partial B_2} = 0 \quad \text{and} \quad \frac{\partial W}{\partial B_3} = 0 \tag{21}$$

From Eqs.(20) and (21) we obtain

$$\sum_i \frac{1}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i}{\sigma_i^2} \times B_2 + \sum_i \frac{z_i}{\sigma_i^2} \times B_3 = \sum_i \frac{y_i}{\sigma_i^2} \tag{22}$$

$$\sum_i \frac{x_i}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i^2}{\sigma_i^2} \times B_2 + \sum_i \frac{x_i z_i}{\sigma_i^2} \times B_3 = \sum_i \frac{x_i y_i}{\sigma_i^2} \tag{23}$$

$$\sum_i \frac{z_i}{\sigma_i^2} \times B_1 + \sum_i \frac{x_i z_i}{\sigma_i^2} \times B_2 + \sum_i \frac{z_i^2}{\sigma_i^2} \times B_3 = \sum_i \frac{z_i y_i}{\sigma_i^2} \tag{24}$$

Again Eqs.(22) - (24) can be written in matrix form as

$$\mathbf{AB} = \mathbf{V} \tag{25}$$

where

$$\mathbf{A} = \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{z_i}{\sigma_i^2} \\ \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{x_i^2}{\sigma_i^2} & \sum_i \frac{x_i z_i}{\sigma_i^2} \\ \sum_i \frac{z_i}{\sigma_i^2} & \sum_i \frac{x_i z_i}{\sigma_i^2} & \sum_i \frac{z_i^2}{\sigma_i^2} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \quad \text{and} \quad \mathbf{V} = \begin{bmatrix} \sum_i \frac{y_i}{\sigma_i^2} \\ \sum_i \frac{x_i y_i}{\sigma_i^2} \\ \sum_i \frac{z_i y_i}{\sigma_i^2} \end{bmatrix} \tag{26}$$

Now let's multiply Eq.(25) by the inverse of matrix **A**

$$\mathbf{A}^{-}\mathbf{A}\mathbf{B} = \mathbf{B} = \mathbf{A}^{-}\mathbf{V}$$

Once again all we have to do to find the fit parameters B_1 , B_2 and B_3, is to invert the **A** matrix and multiply it by the **V** matrix The $\mathbf{A}^{-}$ matrix is also the error matrix with

$$\sigma^2_{B_1} = \mathrm{A}^{-}_{11}, \quad \sigma^2_{B_2} = \mathrm{A}^{-}_{22} \quad \text{and} \quad \sigma^2_{B_3} = \mathrm{A}^{-}_{33}$$

MATRIX INVERSION:

For the various FITS we only have to invert a 2×2 or a 3×3 matrix.

Let's start with the 2×2 matrix.

$$\mathbf{A}^{-}\mathbf{A} = \mathbf{I} \tag{27}$$

Writing Eq.(27) in matrix form

$$\begin{bmatrix} \mathrm{A}^{-}_{11} & \mathrm{A}^{-}_{12} \\ \mathrm{A}^{-}_{21} & \mathrm{A}^{-}_{22} \end{bmatrix} \begin{bmatrix} \mathrm{A}_{11} & \mathrm{A}_{12} \\ \mathrm{A}_{21} & \mathrm{A}_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{28}$$

Doing the matrix multiplication gives

$$\mathrm{A}^{-}_{11}\mathrm{A}_{11} + \mathrm{A}^{-}_{12}\mathrm{A}_{21} = 1 \tag{29}$$

$$\mathrm{A}^{-}_{11}\mathrm{A}_{12} + \mathrm{A}^{-}_{12}\mathrm{A}_{22} = 0 \tag{30}$$

From Eq.(29) and (30)

$$\mathrm{A}^{-}_{11} = \frac{\mathrm{A}_{22}}{\mathbf{D}} \qquad \text{and} \qquad \mathrm{A}^{-}_{12} = -\frac{\mathrm{A}_{12}}{\mathbf{D}}$$

Where the determinant **D** is given by

$$\mathbf{D} = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = A_{11}A_{22} - A_{21}A_{12}$$

Similarly from Eq.(28)

$$\bar{A}_{21}A_{11} + \bar{A}_{22}A_{21} = 0 \tag{31}$$

$$\bar{A}_{21}A_{12} + \bar{A}_{22}A_{22} = 1 \tag{32}$$

Leading to $\bar{A}_{21} = -\dfrac{A_{21}}{\mathbf{D}}$ and $\bar{A}_{22} = \dfrac{A_{11}}{\mathbf{D}}$

Therefore the inverse of **A** is given by

$$\boxed{\mathbf{A}^{-} = \frac{1}{\mathbf{D}}\begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}} \tag{33}$$

Let's now do the 3 x 3 matrix

Generalizing from Eq.(28)

$$\begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ \bar{A}_{21} & \bar{A}_{22} & \bar{A}_{23} \\ \bar{A}_{31} & \bar{A}_{32} & \bar{A}_{33} \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{34}$$

From Eq.(34) we get

$$\begin{aligned} \bar{A}_{11}A_{11} + \bar{A}_{12}A_{21} + \bar{A}_{13}A_{31} &= 1 \\ \bar{A}_{11}A_{12} + \bar{A}_{12}A_{22} + \bar{A}_{13}A_{32} &= 0 \\ \bar{A}_{11}A_{13} + \bar{A}_{12}A_{23} + \bar{A}_{13}A_{33} &= 0 \end{aligned} \tag{35}$$

Solving the set of Eqs.(35) gives

$$\bar{A}_{11} = \frac{A_{22}A_{33} - A_{32}A_{23}}{\mathbf{D}} \qquad \bar{A}_{12} = \frac{A_{12}A_{33} - A_{32}A_{13}}{\mathbf{D}} \qquad \bar{A}_{13} = \frac{A_{12}A_{23} - A_{22}A_{13}}{\mathbf{D}}$$

Where the determinant **D** is given by

$$\mathbf{D} = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = A_{11}(A_{22}A_{33} - A_{32}A_{23}) - A_{21}(A_{12}A_{33} - A_{32}A_{13}) + \cdots etc.$$

Similarly we can find the other six terms, giving the inverse as

$$\mathbf{A}^{-} = \frac{1}{\mathbf{D}} \begin{bmatrix} A_{22}A_{33} - A_{32}A_{23} & A_{32}A_{13} - A_{12}A_{33} & A_{12}A_{23} - A_{22}A_{13} \\ A_{31}A_{23} - A_{21}A_{33} & A_{11}A_{33} - A_{31}A_{13} & A_{21}A_{13} - A_{11}A_{23} \\ A_{21}A_{32} - A_{31}A_{22} & A_{31}A_{12} - A_{11}A_{32} & A_{11}A_{22} - A_{21}A_{12} \end{bmatrix}$$

EXAMPLE: The hyperfine splitting in the ^{57}Fe nucleus seen by the Mossbauer Effect is a good illustration of a fit to two independent variables I_z and I_z^*. The equation describing the splitting is given by

$$\Delta E = \Delta E_{isomeric} + G\, I_z + G^* I_z^* \qquad \text{where} \quad G = g\mu_n B \quad \text{and} \quad G^* = g^*\mu_n B$$

Typical data on the six allowed transitions: ($x_i = \boldsymbol{I_z}$ and $z_i = \boldsymbol{I_z}^*$)

ΔE (mm/s)	σ_i	$\boldsymbol{I_z}$	$\boldsymbol{I_z}^*$
-5.44	2.28×10^{-2}	-0.5	-1.5
-3.20	1.76×10^{-2}	-0.5	-0.5
-0.949	1.45×10^{-2}	-0.5	0.5
0.719	1.45×10^{-2}	0.5	-0.5
2.967	1.76×10^{-2}	0.5	0.5
5.218	2.28×10^{-2}	0.5	1.5

Setting up the **A** matrix of Eq.(26) we take advantage of the symmetries in the data: note that

$$A_{12} = \sum_i \frac{x_i}{\sigma_i^2} = A_{21} = 0 \qquad \text{and} \qquad A_{13} = \sum_i \frac{z_i}{\sigma_i^2} = A_{31} = 0 \qquad \text{where} \quad x_i = \boldsymbol{I_z} \quad \text{and} \quad z_i = \boldsymbol{I_z}^*$$

$$A_{11} = \sum_i \frac{1}{\sigma_i^2} = 19.816\times10^3 \qquad A_{22} = \sum_i \frac{x_i^2}{\sigma_i^2} = 4.954\times10^3 \qquad A_{33} = \sum_i \frac{z_i^2}{\sigma_i^2} = 12.648\times10^3$$

Finally $A_{23} = A_{32} = \sum_i \frac{x_i z_i}{\sigma_i^2} = 2.121 \times 10^3$

$$\mathbf{A} = 10^3 \begin{pmatrix} 19.816 & 0 & 0 \\ 0 & 4.954 & 2.121 \\ 0 & 2.121 & 12.648 \end{pmatrix} \quad \text{and} \quad \mathbf{D} = 10^9 \begin{vmatrix} 19.816 & 0 & 0 \\ 0 & 4.954 & 2.121 \\ 0 & 2.121 & 12.648 \end{vmatrix} = 1.152 \times 10^{12}$$

$$\mathbf{A}^{-} = \frac{10^7}{1.152 \times 10^{12}} \begin{pmatrix} 5.816 & 0 & 0 \\ 0 & 25.063 & -4.203 \\ 0 & -4.203 & 9.816 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = 10^3 \begin{pmatrix} -2.276 \\ 24.17 \\ 36.74 \end{pmatrix}$$

From $\mathbf{B} = \mathbf{A}^{-}\mathbf{V}$ we get

$B_1 = (5.816\text{x}10^7)(-2.276\text{x}10^3)/\ 1.152\text{x}10^{12} = -0.1149$ mm/s

$B_2 = [(25.063)(24.17) + (-4.203)(36.74)]\text{x}10^{10}/1.152\text{x}10^{12} = 3.918$ mm/s

$B_3 = [(-4.203)(24.17) + (9.816)(36.74)]\text{x}10^{10}/1.152\text{x}10^{12} = 2.248$ mm/s

$\sigma_{B_1} = \sqrt{5.816 \times 10^{-5}/1.152} = .0071$

$\sigma_{B_2} = \sqrt{25.063 \times 10^{-5}/1.152} = .0147$

$\sigma_{B_3} = \sqrt{9.816 \times 10^{-5}/1.152} = .0092$

As can be seen from the value of **D** for just six data points this basic approach can quickly lead to overflows for any reasonable number of data points. A more robust method for obtaining the inverse $\mathbf{A}^{-}$ is the Gauss-Jordan elimination method, which avoids the need to calculate the determinant. For a very useful discussion of this method see Numerical Recipes section 2.1 with its strong emphasis on *Pivoting*.